SHARPENING ECONOMIC INTERPRETATION WITH HARS

DAWIS KIM AND TAO ZHA

Abstract. We compare homoskedastic and heteroskedastic specifications under the same economic restrictions in three well-established SVAR applications: monetary policy, oil markets, and fiscal transmission. The heteroskedastic specification yields a credible output contraction after monetary tightening, stronger evidence that oil-specific demand shocks raise oil prices and reduce early economic activity, and a shorter credible spending expansion without a fiscal price puzzle. The differences concern posterior location, response persistence, and interval widths. HARS makes these comparisons feasible through a general MCMC construction whose theoretical foundation combines likelihood-preserving rotations, an explicit posterior target, and posterior-draw repair.

Date: 2nd November 2026.

Key words and phrases. Likelihood preserving rotations, posterior draw repair, shock heteroskedasticity, heavy-tailed fluctuations, 90% credible intervals, monetary policy transmission, oil supply and demand shocks, tax and spending shocks.

JEL classification: C11, C32, E52, E62, Q43.

Kim: University of Virginia; Zha: Federal Reserve Bank of Atlanta, Emory University, and NBER. We thank Andrea Carriero, Lutz Kilian, Eric Leeper, Massimiliano Marcellino, Juan Rubio-Ramirez, and Tommaso Tornese for helpful discussions. Massimiliano Marcellino and Tommaso Tornese kindly provided code implementing their Algorithm 2. The views expressed herein are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the National Bureau of Economic Research.

I. Introduction

Whether monetary tightening contracts output, oil-specific demand shocks raise the real oil price and depress economic activity beyond impact, and government spending shocks raise output without a fiscal price puzzle are central questions in structural vector autoregression (SVAR) analysis. Large movements in output and prices may reflect unusually large shocks, strong dynamic responses, or both. In SVARs, a homoskedastic specification forces tranquil and turbulent episodes to share the same shock variances. This assumption can affect inference about how strongly and how long the economy responds to a shock.

We revisit three widely used applications to ask how their economic conclusions change when shock volatility differs across historical regimes. In the existing literature, these applications use homoskedastic specifications, but their identifying restrictions do not by themselves settle the economic questions above. Under the same identifying restrictions, we compare homoskedastic and heteroskedastic specifications to examine how allowing shock variances to change alters economic inference.

Following Uhlig (2005), a large literature uses sign restrictions to give shocks an economic interpretation. Antolín-Díaz and Rubio-Ramírez (2018, hereafter ADRR) extend this approach with narrative restrictions. We study how accounting for shock volatility changes the probability assigned to their unrestricted dynamic effects. The monetary policy application imposes sign and narrative restrictions tailored to the Volcker episode. The global oil market application uses sign and elasticity restrictions. The fiscal policy application combines sign and narrative restrictions for net-tax and government-spending shocks. In each application, both specifications impose the same economic restrictions. In the heteroskedastic specification, response restrictions must hold for impulse responses with normalized shock volatility and in every variance regime. Narrative restrictions use the volatility prevailing at each historical date. The comparisons therefore reflect the innovation distributions, changing regime variances, and the requirement that the response restrictions hold across regimes.

The monetary policy model follows ADRR's Volcker narrative design, coupled with their traditional sign restrictions. ADRR's central finding is that a contractionary monetary policy shock contracts output. Our homoskedastic specification confirms this finding at the 68% level, but its 90% credible interval contains zero at every horizon. Under the heteroskedastic specification, the 90% interval for real output lies below zero after about two years. Total reserves, whose response is unrestricted, also fall. The economic difference concerns both the output contraction and its transmission through bank reserves.

The oil market application follows Kilian and Murphy (2012, hereafter KM), who identify oil supply, aggregate demand, and oil-specific demand shocks with sign restrictions and a short-run supply-elasticity bound. The 90% oil-price band after an oil-specific demand shock remains above zero through month 7 under the heteroskedastic specification, compared with month 3 under the homoskedastic specification. The early real-activity band remains below

zero through month 3 under the heteroskedastic specification, compared with month 1 under the homoskedastic specification. The price increase and early contraction extend beyond KM's impact restrictions. The heteroskedastic oil-supply price band instead contains zero at every horizon beyond impact. The homoskedastic specification assigns aggregate demand the largest median share in every period. The heteroskedastic specification reveals oil-specific demand's leading role in 1990–2007: its median share of real-oil-price forecast-error variance ranges from 60% to 82%, a majority at all horizons.

For fiscal transmission, we identify net-tax and government-spending shocks with sign and narrative restrictions drawn from major U.S. tax and spending episodes. We normalize each shock to raise its own policy variable by 1% on impact under both specifications. A net-tax increase produces an output contraction credible at the 90% level through quarter 19 under both specifications, consistent with the tax-transmission evidence in Ramey (2016) and Mertens and Ravn (2013). After a spending increase, the heteroskedastic GDP expansion remains credible only through quarter 5, two quarters beyond the restriction window. There is no fiscal price puzzle at the 90% level, and no net-tax increase is credible at any horizon. The homoskedastic specification instead shows a net-tax increase from quarter 3 through quarter 19, accompanied by a GDP expansion through quarter 17 and a price decline at quarters 6–10 at the 90% credible level—a fiscal price puzzle (Jørgensen and Ravn, 2022).

The common finding across these widely used applications is that the heteroskedastic specification can alter economic conclusions under the same identifying restrictions. The differences concern the location and persistence of the responses as well as the width of their 90% credible intervals. Heteroskedastic intervals can be narrower or wider. What matters is how changes in posterior probabilities alter economic inference at the conventional 90% credible level.

Our heteroskedastic specification follows Brunnermeier et al. (2021, hereafter BPSS). It lets large disturbances be large without forcing every period to share the same shock variances. The BPSS likelihood also allows for heavy tails. But heteroskedasticity does not, by itself, solve economic identification. The likelihood can separate model shocks. It does not tell us which shock is monetary policy, oil supply, oil-specific demand, net taxes, or government spending. This labeling problem is a weakness of identification through heteroskedasticity, a concern emphasized by Herwartz and Lütkepohl (2014). Heteroskedasticity alone leaves these labels to ex post inspection.

Following Uhlig (2005) and Antolín-Díaz and Rubio-Ramírez (2018), we let economic identification come from sign and narrative restrictions on rotations of standardized shocks. Combining regime-dependent volatility with these restrictions presents a technical challenge. An unabsorbed coefficient rotation mixes shocks whose variances differ across regimes. The reduced-form coefficients remain unchanged, but the covariance matrices and the likelihood generally change. Our key construction absorbs regime volatility into the coefficient matrices before rotation. This construction transforms the shocks to unit variance in every regime.

We then rotate the entire absorbed block. The rotation changes the impulse responses and historical shocks. It leaves the heteroskedastic likelihood unchanged. We call the resulting procedure the heteroskedastic absorb-and-rotate sampler (HARS). HARS combines posterior simulation of model parameters with rotations that satisfy the identifying restrictions. The theoretical foundation that makes this combination valid is the paper's central methodological contribution.

In HARS, the likelihood and the economic restrictions have distinct roles. The BPSS likelihood captures heavy-tailed fluctuations and changes in shock volatility across regimes. Economic restrictions on the rotation provide identification. Neither part substitutes for the other. Economic restrictions do not make a homoskedastic likelihood account for changing shock variances. Heteroskedasticity alone does not ensure that the implied responses have the intended economic meaning. HARS makes the two parts work together. HARS is a general construction for adapting standard Markov chain Monte Carlo (MCMC) posterior simulators for the heteroskedastic model, including Gibbs and Metropolis–Hastings samplers. We derive the admissibility verifications and acceptance probabilities needed to make their updates preserve the restricted joint posterior. The construction is not tied to the BPSS posterior simulator; it works with other MCMC simulators through the target-preserving updates developed in the paper. These updates make the empirical comparisons feasible.

Two closed-form analytical examples with two variables, no lags, and two regimes explain how economic restrictions can change response probabilities even when the impulse response formula is identical across models. The first example isolates sign verifications across regimes. The second isolates historical shock recovery under a narrative restriction. The model parameters are fixed; the admissible rotations are random. Posterior response probabilities average these conditional probabilities over the parameter marginal of the restricted joint target. The closed-form solutions establish how inference can change. They help explain how the same economic restrictions can lead the heteroskedastic and homoskedastic specifications to assign different posterior probabilities to impulse responses in the three empirical applications.

The paper makes three contributions. First, it shows how the heteroskedastic specification alters the economic implications of three established SVAR applications. Second, it provides transparent examples and an explicit posterior target that explain the role of the identifying restrictions. Third, it develops a general likelihood-preserving MCMC construction and a practical HARS implementation, including posterior-draw repair when admissible completions are rare. The methodology makes the empirical comparisons between heteroskedastic and homoskedastic specifications feasible.

The paper is related to the literature on time-varying VARs. Primiceri (2005) allows coefficients and shock volatility to change over time. Changes in shock volatility have also been studied in Markov-switching models that estimate latent regimes and their transition probabilities (Sims and Zha, 2006; Bianchi et al., 2022). BPSS choose variance-regime dates

using observed changes in the data and historical knowledge. They argue that conditioning on these dates should have little effect on inference about model dynamics when regimes are few, persistent, and clearly indicated by the data. We follow BPSS and condition on fixed variance-regime dates while keeping the propagation coefficients fixed across regimes.

Our study focuses on how changing shock volatility affects inference within established empirical specifications. For a transparent comparison, we retain each application's variables, lags, and economic restrictions and follow existing empirical specifications in choosing fixed dates for variance changes. The heteroskedastic specification lets shock variances change across regimes. It also has the Student-t shocks, while the homoskedastic specifications have Gaussian shocks. The comparisons do not establish that either specification is generally superior. New applications with estimated Markov-switching regimes are beyond the scope of this paper and are left for future research. Even with fixed regime dates, imposing sign and narrative restrictions while preserving the heteroskedastic likelihood is a very difficult methodological problem.

Arias et al. (2026c) develop inference for time-varying SVARs identified with sign restrictions. Their Proposition 2 establishes an important difficulty: BPSS models fall outside their class of rotation-invariant time-varying SVARs, which requires a rotation-invariant structural prior. We address the difficulty of imposing economic restrictions in BPSS models through a different construction. We absorb regime volatility before rotation, preserve the heteroskedastic likelihood, and define the joint posterior target for the model parameters and admissible rotations.

Carriero et al. (2024, hereafter CMT) combine identifying information from heteroskedasticity with sign, narrative, and proxy restrictions. Their framework also allows dated variance regimes, as in BPSS. Their Algorithm 2 and its narrative extension impose the restrictions by direct accept–reject on model-parameter draws. HARS instead imposes the economic restrictions on rotations of standardized shocks and updates the model parameters for the corresponding joint target. The methods therefore define different posterior objects. Direct accept–reject becomes costly when few draws satisfy the restrictions. Section VIII compares the computational implementations under their respective targets.

The rest of the paper proceeds as follows. Section II presents the heteroskedastic model. Sections III and IV develop the target distribution and sampler, and Section V presents HARS. Section VI gives the two analytical examples. Section VII studies monetary policy, oil markets, and fiscal policy. Section VIII examines computational efficiency, and Section IX concludes. The appendix contains proofs, complete example derivations, data sources, and computational details.

II. The Model

As discussed in the introduction, we use BPSS's heteroskedastic specification to capture large movements and fat-tailed distributions in macroeconomic data. We impose sign and

narrative restrictions for economic identification. These restrictions act on rotations of standardized shocks, constructed to preserve the heteroskedastic likelihood and the dynamics of the data.

II.1. **The heteroskedastic model.** Let y_t be an $n \times 1$ vector of observables and x_t the $r \times 1$ vector of predetermined terms, including lagged observables and the constant. Following the deterministic variance-regime specification in BPSS, let

$$s_t : \{1, \ldots, T\} \to \{1, \ldots, h\}$$

be a known date-to-regime map. The regime associated with each date is treated as fixed and observed for statistical inference.

The SVAR setup is

$$y_t' A_0 = x_t' A_+ + \eta_t', \tag{1}$$

where A_0 is $n \times n$ and nonsingular, A_+ is $r \times n$, and $E(\eta_t \eta_t') = \Lambda_{s_t}$ with

$$\Lambda_k = \mathrm{diag}(\lambda_{1,k}, \ldots, \lambda_{n,k}),\ \lambda_{i,k} > 0,\ i = 1, \ldots, n,\ k = 1, \ldots, h.$$

Conditional on regime s_t, the components of the shock vector η_t are independent across equations and time.

BPSS allow heavy-tailed shocks through a normal mixture specification:

$$\eta_{i,t} \mid \nu_{i,t}, s_t = k \sim \mathcal{N}(0, \lambda_{i,k}\nu_{i,t}),$$

where $\nu_t = (\nu_{1,t}, \ldots, \nu_{n,t})'$ is a vector of latent variance multipliers, normalized to have mean one. Thus, Λ_k retains its interpretation as the regime-dependent shock variance matrix, following BPSS (p. 1852, fn. 11). BPSS include ν_t as a latent block in their posterior simulator. Our analysis follows BPSS and uses Student-t shocks with 5.7 degrees of freedom within each variance regime. The rotation, sign admissibility, and impulse response mapping depend on (A_0, Λ, A_+) and Q, not on ν_t. We therefore suppress ν_t throughout the remainder of the paper.

BPSS impose the variance normalization

$$\frac{1}{h} \sum_{k=1}^{h} \lambda_{i,k} = 1, \qquad i = 1, \ldots, n,$$

which fixes the scale of each shock variance path across regimes. The diagonal entries of $\Lambda_{s_t}^{1/2}$ are the shock standard deviations in the regime assigned to date t. Heteroskedasticity enters only through this deterministic regime-dependent diagonal scaling.

In this paper, we use the BPSS regime-heteroskedastic framework as a likelihood-based and computationally efficient way to accommodate regime-dependent volatility and heavy-tailed fluctuations in the data. We do not treat the shocks η_t as economically interpreted objects. At this stage they are model shocks: independent across equations and time, with

variances that change across regimes. The economic interpretation comes later, through an orthogonal rotation Q selected by sign and narrative restrictions.

This separation is essential. The rotation Q is not meant to change the dynamics of the data. Its role is to select an admissible economic interpretation while leaving the reduced-form data-generating process unchanged. This invariance is a fundamental principle in the sign-restriction literature, and we preserve it. The difficulty is that the rotation commonly used in homoskedastic SVARs preserves the reduced form but need not preserve diagonal shock covariance once shock variances differ across regimes. The following discussion makes this point explicit.

The reduced form follows by post-multiplying (1) by A_0^{-1}:

$$y_t' = x_t'B + u_t', \qquad B = A_+A_0^{-1}, \qquad u_t' = \eta_t'A_0^{-1}.$$

The residual covariance is therefore regime-dependent:

$$\Sigma_{s_t} = \mathrm{Var}(u_t \mid s_t) = (A_0^{-1})'\Lambda_{s_t}A_0^{-1}. \tag{2}$$

In regime k, the compact form is[1]

$$\Sigma_k = (A_0^{-1})'\Lambda_k A_0^{-1}.$$

The reduced-form coefficient matrix remains invariant under the coefficient-only rotation

$$(A_0, A_+) \mapsto (A_0Q, A_+Q).$$

In the Gaussian homoskedastic benchmark, this rotation preserves the shock law. Under heteroskedasticity, preserving the full data distribution also requires transforming that law. If the shock vector is transformed along with the coefficients, the algebra remains unchanged because

$$y_t'A_0Q = x_t'A_+Q + \eta_t'Q$$

solves back to

$$y_t' = x_t'B + \eta_t'QQ'A_0^{-1} = x_t'B + \eta_t'A_0^{-1}.$$

The transformed shock is now $\eta_t'Q$. Under heteroskedasticity, η_t has diagonal covariance Λ_{s_t}, not scalar covariance. A commonly used orthogonal rotation then mixes shocks with different variances. The next lemma shows when this rotation preserves diagonal shock covariance.

Lemma 1. *Fix a regime k and let $Q \in \mathcal{O}(n)$. Define the rotated shock*

$$\widetilde{\eta}' = \eta'Q.$$

The components of $\widetilde{\eta}$ are mutually uncorrelated if and only if

$$Q'\Lambda_k Q$$

[1]The transpose placement in (2) follows from the right-multiplication convention $y_t'A_0$ in (1), under which columns of A_0 correspond to equations. Under the alternative convention A_0y_t, the corresponding covariance expression is $A_0^{-1}\Lambda_k(A_0^{-1})'$.

is diagonal. If the diagonal entries of Λ_k are distinct, this condition holds only for signed permutation matrices. More generally, up to a permutation of its columns, Q may rotate only within equal-variance blocks. Under Gaussian shocks, uncorrelatedness also implies independence; with scale mixtures, it need not.

Proof. See Appendix S1.1. □

Lemma 1 explains why the usual covariance argument for homoskedastic SVARs cannot be carried over unchanged. A coefficient rotation mixes shocks with different regime variances and generally destroys diagonal shock covariance. When Λ_k is proportional to the identity, every orthogonal rotation preserves the covariance. Under heteroskedasticity, the model must first return the shocks to unit variance and only then rotate. The absorbed form presented in the next section does exactly that. For heavy-tailed shocks, the full joint shock density must also be transformed, as described below.

II.2. **Absorbed form and impulse responses.** The regime variances are absorbed into the coefficient matrices so that the shocks have unit variance. Let

$$\Xi_{s_t} = \Lambda_{s_t}^{-1/2}.$$

Post-multiplying (1) by Ξ_{s_t} gives

$$y_t' A_0 \Xi_{s_t} = x_t' A_+ \Xi_{s_t} + \varepsilon_t', \tag{3}$$

where

$$\varepsilon_t = \Xi_{s_t} \eta_t, \qquad E(\varepsilon_t \varepsilon_t' \mid s_t) = I_n.$$

Equation (3) is algebraically identical to (1). The shocks now have unit variance in every regime, so an orthogonal rotation Q preserves their unit-variance covariance matrix. The regime-specific information is absorbed into the block $(A_0 \Xi_{s_t}, A_+ \Xi_{s_t})$.

The identifying restrictions enter through an orthogonal rotation Q, following Rubio-Ramírez et al. (2010). Post-multiplying the entire absorbed block in (3) by an orthogonal matrix $Q \in \mathcal{O}(n)$ yields

$$y_t' A_0 \Xi_{s_t} Q = x_t' A_+ \Xi_{s_t} Q + \varepsilon_t' Q, \tag{4}$$

where the column vector of rotated shocks is $e_t = Q' \varepsilon_t$. These shocks have unit variance and are mutually uncorrelated. Under the scale-mixture specification, they need not be independent: their density is the density induced by this transformation, not a new product of independent shock densities. Conditional on the latent scales, their covariance is $Q' \operatorname{diag}(\nu_t) Q$. This induced joint distribution is used throughout the likelihood argument. Let $\{\Phi_\ell(B)\}_{\ell \geq 0}$ denote the moving-average coefficients for the column-vector reduced form $y_t = B' x_t + u_t$, with $\Phi_0(B) = I_n$.[2] Solving (4) forward, the dynamic responses of the observables to rotated

[2]We index the impulse-response horizon by ℓ to keep it distinct from the regime index k.

unit-variance shocks in regime k at horizon ℓ are

$$\Theta_{k,\ell}(Q) = \Phi_\ell(B)\,(A_0^{-1})'\Lambda_k^{1/2}Q, \tag{5}$$

because

$$\left((A_0\Xi_k Q)^{-1}\right)' = (A_0^{-1})'\Lambda_k^{1/2}Q$$

and the reduced-form coefficient remains B. Rows of $\Theta_{k,\ell}(Q)$ index observables and columns index rotated shocks. The regime enters through $\Lambda_k^{1/2}$ before the rotation; it does not generally rescale the columns of the rotated impact matrix separately. We report impulse responses at the BPSS average scale $\Lambda_k = I_n$, so that $\Theta_\ell(Q) = \Phi_\ell(B)\,(A_0^{-1})'Q$. Sign and elasticity restrictions are imposed on the responses at this scale and on the regime responses $\Theta_{k,\ell}(Q)$ for every $k = 1, \ldots, h$, with the same orientation of each shock throughout: the same economic shock must satisfy the same restrictions in every regime. Historical shocks and narrative restrictions use the variance of the regime prevailing at each date.

III. Theoretical Results

Let Y denote the sample data and let $\theta = (A_0, \Lambda, A_+)$, where $\Lambda = \{\Lambda_k\}_{k=1}^h$. The posterior density for θ is proportional to the heteroskedastic likelihood times the prior density for θ. The BPSS prior and Markov chain Monte Carlo (MCMC) sampler therefore deliver posterior draws of the heteroskedastic system. Sign and narrative restrictions provide identification by selecting admissible rotations of those draws.

Equation (5) shows that the rotation Q changes the dynamic responses associated with a posterior draw. It changes the impact responses and, through the moving-average coefficients $\{\Phi_\ell(B)\}_{\ell\geq 0}$, the responses at all subsequent horizons. But Q does not reweight the draw through the likelihood. Given the regime path, the likelihood depends on (A_0, Λ, A_+), not on Q. The rotation selects an admissible economic interpretation of the responses generated by a posterior draw; it does not change the heteroskedastic likelihood assigned to that draw. The following proposition formalizes this invariance. This distinction is what makes sign-restricted inference under heteroskedasticity coherent.

Proposition 1. *Fix θ and the regime path. For every $Q \in \mathcal{O}(n)$, the absorbed rotation leaves the likelihood unchanged:*

$$p(Y \mid \theta, Q) = p(Y \mid \theta), \tag{6}$$

where the left-hand side denotes the likelihood implied by (4).

Proof. See Appendix S1.2. □

Proposition 1 establishes likelihood invariance for rotations of standardized shocks. HARS, presented in Section V, includes a continuous rotation block and targets the joint posterior in Definition 1. The rotation is sampled from its restricted-Haar conditional, and the model parameters are updated consistently with the same restrictions.

The likelihood invariance established above follows from our key construction: absorbing regime volatility before rotation. The next proposition makes this difficulty precise at the likelihood level: rotating the original coefficient matrices while holding regime variances fixed generally fails to preserve the likelihood.

Proposition 2. *Fix Λ and rotate only the coefficients, $(A_0, A_+) \mapsto (A_0Q, A_+Q)$, while retaining the original shock law. Likelihood invariance for all data requires*

$$Q\Lambda_k Q' = \Lambda_k$$

for every k. In particular, likelihood invariance fails generically whenever some Λ_k has unequal diagonal entries.

Proof. See Appendix S1.3. □

Proposition 2 is the crux. Under homoskedasticity, $\Lambda_k \propto I_n$ and thus Q cancels from the covariance. With Gaussian shocks this also preserves the likelihood. Under heteroskedasticity, the same coefficient-only rotation generally changes even the covariance and therefore the likelihood. A rotation used to identify structural shocks instead refits the model. The absorbed rotation avoids this tilt. It leaves the heteroskedastic likelihood unchanged and uses Q only to search for an admissible economic interpretation.

The vast literature utilizes accept–reject sampling to draw the rotation matrix Q. Under heteroskedasticity, the critical question is which joint distribution of (θ, Q) the restrictions define, and how a sampler can target it. We state the target explicitly.

Definition 1. Let $\theta = (A_0, \Lambda, A_+)$, $p(\theta \mid Y)$ be the heteroskedastic posterior density, and μ be Haar probability measure on $\mathcal{O}(n)$. Let $SR(\theta, Q)$ be the indicator that the sign and narrative restrictions are satisfied:

$$SR(\theta, Q) \equiv \mathbf{1}\{(\theta, Q) \text{ satisfies the sign and narrative restrictions}\}.$$

Define

$$R(\theta) = \{Q \in \mathcal{O}(n) : SR(\theta, Q) > 0\}, \quad v(\theta) = \mu(R(\theta)) = \Pr(SR = 1 \mid \theta, Y),$$

and assume $c(Y) = \int p(\theta \mid Y)\, v(\theta)\, d\theta > 0$. The sign-restricted posterior target is

$$\pi(d\theta, dQ \mid Y, SR) = \frac{p(\theta \mid Y)\, \mathbf{1}\{Q \in R(\theta)\}}{c(Y)}\, d\theta\, \mu(dQ).$$

Definition 1 imposes the restrictions on the joint distribution of (θ, Q). Before the restrictions are imposed, θ has the heteroskedastic posterior and Q has Haar measure; by Proposition 1, the data do not update Q. With sign restrictions alone, Definition 1 is therefore the posterior implied by the BPSS prior on θ and Haar measure on Q, truncated to the admissible region, as in Arias et al. (2018) for homoskedastic SVARs. Narrative restrictions also involve the data at the narrative dates, and Definition 1 is then conditional on that narrative record.

The target has three implications, stated formally in Proposition S1 of Appendix S1.4. Conditional on θ, the rotation Q has Haar measure restricted to $R(\theta)$. Conditional on Q, the model parameters have density proportional to $p(\theta \mid Y)\,\mathbf{1}\{SR(\theta, Q) > 0\}$. The marginal density of θ is proportional to $p(\theta \mid Y)\,v(\theta)$. The restrictions therefore affect the model parameters only through the Haar volume of their admissible rotations. This weight does not come from the likelihood, which Proposition 1 leaves unchanged, and it excludes no parameter value with positive admissible volume. Nor does it need to be computed: a sampler that alternates updates from the two conditional distributions accounts for $v(\theta)$ through its admissibility checks. HARS is such a sampler. It updates Q given θ by restricted-Haar sampling (Section IV.1) and θ given Q by completed-draw updates with posterior-draw repair (Section IV.2).

IV. Sign and Narrative Restrictions

Section III establishes that the rotation Q selects an economic interpretation of a retained heteroskedastic draw without entering the likelihood. This section turns that principle into a sampler, isolating three pitfalls where a naive implementation silently alters the target. The first involves how an admissible rotation is drawn. The second addresses how a posterior move is repaired when it renders the current rotation inadmissible. The third clarifies the valid applications of rejection counts. The repair result is stated as a theorem in the main text; the supporting results on drawing rotations and on rejection counts are stated and proved in Appendix S1.

Our construction uses one computational tool from Arias et al. (2026b) for a different inferential problem. They utilize the elliptical slice sampling (ESS) method of Murray et al. (2010) to sample admissible rotations in homoskedastic SVARs with sign restrictions. HARS addresses heteroskedastic SVARs, where ordinary coefficient rotations fail to preserve the likelihood. Our key construction absorbs regime volatility before rotation and preserves the heteroskedastic likelihood. ESS supplies only the rotation updates. HARS builds on a standard MCMC posterior simulator for the model parameters and adds posterior-draw repair when admissible completions are rare. The parameter and rotation updates preserve the same explicit joint posterior target. Our contribution is the likelihood-preserving construction and the joint sampler that makes sign and narrative restrictions feasible in these models.

IV.1. **Drawing an admissible rotation.**

For a retained draw θ, Definition 1 gives the conditional target of the rotation: Haar measure restricted to the admissible set $R(\theta)$. The rotation step samples only over admissible rotations for the retained draw. It does not redraw, reject, or reweight θ.

Let X be an $n \times n$ matrix with independent standard normal entries. Let $\gamma(X)$ denote the QR map that normalizes the R factor to a positive diagonal and returns the orthogonal factor.

This is the Gaussian-to-Haar construction of Rubio-Ramírez et al. (2010): $\gamma(X)$ is Haar distributed on $\mathcal{O}(n)$. Restricting the Gaussian distribution of X to the event $SR(\theta, \gamma(X)) > 0$ therefore yields a distribution whose image under γ is the restricted-Haar target. The full conditional density of each column of X is a standard Gaussian density times an admissibility indicator, so a column-wise ESS update leaves it invariant. Proposition S2 in Appendix S1.5 states and proves these properties. HARS carries the Gaussian matrix X from one iteration to the next, and the parameter update conditions on the current rotation $Q = \gamma(X)$. Under heteroskedasticity, rotating the unabsorbed coefficients generally changes the likelihood (Proposition 2). The absorbed rotation is likelihood-preserving (Proposition 1). This construction allows us to update admissible rotations in the heteroskedastic model using the restricted-Haar conditional in Definition 1.

IV.2. **Repairing a stale rotation.** A rotation that is admissible for the current draw of θ can be inadmissible for a proposed new draw. We refer to this event as staleness:

$$SR(\theta^\star, Q) \leq 0,$$

where $\theta^\star = (A_0^\star, \Lambda^\star, A_+^\star)$ is the proposed parameter draw and Q is the current rotation. The event can follow a move in (A_0, Λ), because the restricted responses include the regime responses, which depend on $\Lambda_k^{1/2}$. It can also follow a redraw of A_+ when response restrictions bind beyond impact or narrative restrictions depend on recovered historical shocks. The repair must restore admissibility while leaving the conditional target of the completed draw invariant. The next theorem shows that the obvious shortcut fails this invariance requirement and gives a repair that satisfies it.

Partition the parameter collection and the conditioning set as

$$z_0 = (A_0, \Lambda), \quad z_+ = A_+, \quad \theta = (z_0, z_+),$$

and let s_0 collect everything held fixed in the block update: the data, the regime path, the prior hyperparameters, the latent scales, and any other conditioning objects. Let $p_0(z_0 \mid s_0)$ be the collapsed density of the first block with A_+ integrated out, and let $p_+(z_+ \mid z_0, s_0)$ be the Gibbs conditional density of the companion block.

Theorem 1 (A stale rotation must be repaired on the completed draw)**.** *Fix Q and s_0, and assume that the admissible region has positive probability under $p_0 p_+$. By Definition 1, the sign-restricted target of the completed draw is*

$$\pi^{SR}(z_0, z_+ \mid s_0, Q) \propto p_0(z_0 \mid s_0)\, p_+(z_+ \mid z_0, s_0)\, \mathbf{1}\{SR(z_0, z_+, Q) > 0\}.$$

Define the admissibility weight

$$\omega_0(z_0; s_0, Q) = \int p_+(u \mid z_0, s_0)\, \mathbf{1}\{SR(z_0, u, Q) > 0\}\, du,$$

the probability that a Gibbs draw of A_+ completes z_0 into an admissible draw. Then the sign-restricted marginal density of z_0 is proportional to

$$p_0(z_0 \mid s_0)\,\omega_0(z_0; s_0, Q).$$

Consequently:

(i) Assume $\omega_0(z_0; s_0, Q) > 0$ for $p_0(\cdot \mid s_0)$-almost every z_0. Suppose z_0 is drawn from $p_0(\cdot \mid s_0)$ and, holding it fixed, only $z_+ \sim p_+(\cdot \mid z_0, s_0)$ is redrawn until $SR(z_0, z_+, Q) > 0$. This is a valid conditional *update for z_+, but the resulting marginal density of z_0 is $p_0(z_0 \mid s_0)$. The target marginal is proportional to $p_0(z_0 \mid s_0)\,\omega_0(z_0; s_0, Q)$. These two marginals coincide only when $\omega_0(z_0; s_0, Q)$ is constant for p_0-almost every z_0.*

(ii) Let $q_0(\cdot \mid a)$ be the first-block proposal density, with ordinary MH acceptance probability

$$\alpha_{\mathrm{MH}}(a, b) = \min\left\{1,\ \frac{p_0(b \mid s_0) q_0(a \mid b)}{p_0(a \mid s_0) q_0(b \mid a)}\right\}.$$

Let K_0 be this first-block MH transition, including its rejection outcome. It is reversible with respect to $p_0(\cdot \mid s_0)$. Hold Q, s_0, and the proposal settings fixed throughout the update.

From an admissible current draw $\theta = (z_0, z_+)$, draw $z_0^\star$ by one transition of K_0 from z_0, then draw $z_+^\star \sim p_+(\cdot \mid z_0^\star, s_0)$, even if the first-block MH test rejects. Set $\theta^\star = (z_0^\star, z_+^\star)$. If it is admissible, retain it without another MH test.

Otherwise, for a fixed integer $N_\theta \geq 0$, make at most N_θ additional completed transitions. At repair attempt $\nu = 1, \ldots, N_\theta$, draw

$$z_{0,\nu}^{\mathrm{cand}} \sim K_0(z_0^\star, \cdot), \qquad z_{+,\nu}^{\mathrm{cand}} \sim p_+(\cdot \mid z_{0,\nu}^{\mathrm{cand}}, s_0),$$

and set $\theta_\nu^{\mathrm{cand}} = (z_{0,\nu}^{\mathrm{cand}}, z_{+,\nu}^{\mathrm{cand}})$. Retain the first admissible candidate with probability one and stop. After a failed check, set $\theta^\star \leftarrow \theta_\nu^{\mathrm{cand}}$ before the next transition. If the budget is exhausted without an admissible return, retain both components of the original admissible θ. When $N_\theta = 0$, make no additional transition. For every fixed total budget $N = N_\theta + 1 \geq 1$, this transition is reversible with respect to $\pi^{SR}(\cdot \mid s_0, Q)$ and leaves it invariant. No additional endpoint MH test or failed-history product is required.

Proof. See Appendix S1.6. □

Theorem 1 is the foundation of the joint repair used in Section V. Drawing A_+ from $p(z_+ \mid z_0, s_0)$ is computationally inexpensive. Therefore, it is tempting to try A_+ until it works. But this naive procedure is incorrect. First-block values for which admissible completions are rare should receive less posterior mass, and only the completed-draw repair enforces that. The conditional redraw of A_+ is legitimate in exactly one situation: as a companion-block update when z_0 already carries its correct sign-restricted marginal target. It is not a legitimate route to rescue a completed draw that has just failed.

The boundary case is instructive. When $\omega_0(z_0; s_0, Q) = 0$, the first block z_0 has no admissible completion, and the target assigns it zero mass. The completed-draw repair enforces this on its own: it never retains such a candidate. The conditional redraw of $z_+ \equiv A_+$ does not. Holding that z_0 fixed, it searches for an admissible A_+ that does not exist and never terminates. The shortcut therefore fails exactly where the target is most decisive.

The direction of the error is also clear. The conditional redraw of A_+ removes the factor $\omega_0(z_0; s_0, Q)$, so it inflates first-block values with small ω_0 relative to the truth. These are the values whose sign-admissible completions are scarce. Then the posterior distribution over (A_0, Λ), and hence over the impulse responses, tilts toward the parameter configurations the restrictions are designed to make rare. The error therefore favors the very structures the identification was designed to discipline.

Part (ii) of Theorem 1 uses first-block MH transitions followed by companion draws. With $N = 1$, it makes one completed transition, retains the candidate if it is admissible, and otherwise retains the original pair. For $N > 1$, each additional transition starts from the last failed candidate. Reversibility of the completed transition permits the first admissible return to be retained without another MH test. The repair requires neither $v(\theta)$ nor ω_0. A repair that finds no admissible candidate within the fixed budget retains the original admissible pair. That outcome is part of the transition, not a discarded iteration.

IV.3. **Rejection counts are diagnostics, not density terms.** Rotation searches and repair attempts produce rejection counts. It is tempting to read a high rejection rate as evidence that a draw fits the data poorly, or to fold it into a likelihood or posterior calculation. Both interpretations are incorrect. For a fixed θ, the first admissible rotation among independent Haar proposals has the restricted-Haar distribution of Definition 1, and the likelihood of θ is unchanged. The admissible volume $v(\theta)$ does enter the marginal of θ in Definition 1, but the parameter update accounts for it through its admissibility checks. Realized rejection counts depend on the conditioning state and on the proposal mechanism, so they must not be used as additional weights. Proposition S3 in Appendix S1.7 states these results formally.

Failed candidates are auxiliary proposals. They are not retained draws of the model parameters and do not enter posterior averages. A frequent need to repair is a diagnostic of a small admissible set, or of a rotation that often becomes stale after updates to θ. It does not imply that the model fits the data poorly. Section V combines these results into a single sampler.

V. Algorithm

The sampler alternates an update of the model parameters given the current rotation with an update of the rotation given the model parameters. The first block $z_0 = (A_0, \Lambda)$ is

moved by MH on the completed draw, with posterior-draw repair when a proposal makes the current rotation inadmissible (Theorem 1). The companion block $z_+ = A_+$ is refreshed from its Gaussian Gibbs conditional, subject to the same admissibility check. Latent scales and other objects collected in s_0 are updated by the BPSS sampler; they do not enter the restrictions. The rotation is then updated by the column-wise ESS step of Proposition S2, with the model parameters held fixed.

The parameter update must condition on the rotation because admissibility depends on the model parameters. Restrictions on impulse responses are evaluated at the BPSS average scale, where the impact matrix, with observables in rows, is $(A_0^{-1})'Q$, and in every regime k, where it is $S_k = (A_0^{-1})'\Lambda_k^{1/2}Q$. A move in Λ_k changes the relative weights inside S_k and can make an admissible rotation inadmissible. Restrictions beyond impact also depend on A_+ through B, and narrative restrictions depend on Λ through historical shocks standardized by the variance of the regime prevailing at each narrative date. Sign restrictions are imposed on impulse responses and historical shocks, never on Λ as an object. Admissibility must therefore be checked for every proposed move in θ. When response restrictions bind only at impact and no narrative restriction depends on recovered shocks, A_+ does not enter the admissibility check. Throughout, the admissible set is

$$R(\theta) = \{Q \in \mathcal{O}(n) : SR(\theta, Q) > 0\}.$$

There are two tuning parameters:

- $N_{\text{ess}} \geq 1$, the number of ESS updates of Q per iteration;
- $N_\theta \geq 0$, the maximum number of additional completed candidates in a posterior-draw repair.

For initialization, use the column-by-column construction of Arias et al. (2026b, Algorithm 4) to propose a complete rotation. Build $Q = [q_1, \ldots, q_n]$ one column at a time, projecting each Gaussian draw onto the orthogonal complement of the columns already fixed, renormalizing, and checking its sign restrictions. Then check the full indicator $SR(\theta, Q)$, including all sign, elasticity, and narrative restrictions. Accept the initial pair only if $SR(\theta, Q) = 1$; otherwise repeat the initialization with another candidate pair (θ, Q). Set $X = Q$ only after this full check. This gives $\gamma(X) = Q$ under the positive-diagonal convention and seeds the ESS step of Proposition S2. Burn-in draws are discarded as usual. The retained admissible parameter vector θ remains available throughout the repair. The symbol $\theta^\star$ denotes a working completed candidate; it can fail the restrictions and is not then a retained posterior draw.

Algorithm 1. *Start from (θ, Q) with $SR(\theta, Q) = 1$, and carry the corresponding ESS matrix X.*

Step 1: Posterior draw with admissibility check. *Propose $z_0^{\text{prop}} \sim q_0(\cdot \mid z_0)$. Accept this first-block proposal with probability $\alpha_{\text{MH}}(z_0, z_0^{\text{prop}})$. Set $z_0^\star = z_0^{\text{prop}}$ on acceptance and*

$z_0^\star = z_0$ *on rejection. Then draw* $z_+^\star \sim p_+(\cdot \mid z_0^\star, s_0)$ *and form* $\theta^\star = (z_0^\star, z_+^\star)$. *The companion is drawn even if the first-block MH test rejects.*

- *If* $SR(\theta^\star, Q) = 1$, *set* $\theta \leftarrow \theta^\star$ *and proceed to Step 3. There is no additional MH test.*
- *If* $SR(\theta^\star, Q) = 0$, *when* $N_\theta = 0$, *proceed to Step 3 with the original admissible* θ; *when* $N_\theta > 0$, *keep the original admissible* θ *unchanged and* $\theta^\star$ *as the working failed candidate, then proceed to Step 2.*

Step 2: Posterior-draw repair from the last failed candidate. *For* $\nu = 1, \ldots, N_\theta$:

(a) *Propose* $z_0^{\text{prop}} \sim q_0(\cdot \mid z_0^\star)$. *Make the ordinary first-block MH decision with probability*

$$\alpha_{\text{MH}}(z_0^\star, z_0^{\text{prop}}).$$

Set $z_{0,\nu}^{\text{cand}} = z_0^{\text{prop}}$ *on acceptance and* $z_{0,\nu}^{\text{cand}} = z_0^\star$ *on rejection.*

(b) *Draw* $z_{+,\nu}^{\text{cand}} \sim p_+(\cdot \mid z_{0,\nu}^{\text{cand}}, s_0)$ *and form* $\theta_\nu^{\text{cand}} = (z_{0,\nu}^{\text{cand}}, z_{+,\nu}^{\text{cand}})$. *Draw this companion even when the first-block MH proposal has been rejected.*

(c) *If* $SR(\theta_\nu^{\text{cand}}, Q) = 1$, *set* $\theta \leftarrow \theta_\nu^{\text{cand}}$, *stop the repair, and proceed to Step 3. Retain this first admissible return with probability one; there is no extra endpoint MH test or failed-history product.*

(d) *If* $SR(\theta_\nu^{\text{cand}}, Q) = 0$, *set* $\theta^\star \leftarrow \theta_\nu^{\text{cand}}$ *and continue to the next* ν. *The next first-block proposal therefore starts from this last failed candidate.*

Failed candidates generate the next repair transition but are never saved as posterior draws. If no candidate is admissible when ν *reaches the budget* N_θ, *discard the auxiliary path and retain both components of the original admissible* θ. *Proceed to Step 3. Do not retain the last inadmissible candidate after budget exhaustion.*

Step 3: Companion refresh and latent updates. *Hold the retained* z_0 *fixed and draw* $z_+^{\text{prop}} \sim p_+(\cdot \mid z_0, s_0)$. *Set* $z_+ \leftarrow z_+^{\text{prop}}$ *if* $SR(z_0, z_+^{\text{prop}}, Q) = 1$; *otherwise retain the current* z_+. *Then update latent variables and other objects absent from SR using their BPSS full conditionals. These updates use the retained admissible* θ, *not a failed working candidate.*

Step 4: Rotation draw. *Hold the retained* θ *fixed. Starting from the carried* X, *perform* N_{ess} *column-wise ESS updates to obtain* $X^{(r)}$ *and* $Q^{(r)} = \gamma(X^{(r)})$, $r = 1, \ldots, N_{\text{ess}}$. *Each satisfies* $SR(\theta, Q^{(r)}) = 1$. *Save* $(\theta, Q^{(r)})$ *for impulse-response calculations.*

Step 5: Advance the state. *Set* $X \leftarrow X^{(N_{\text{ess}})}$ *and* $Q \leftarrow Q^{(N_{\text{ess}})}$, *retain* θ, *and return to Step 1.*

Section IV shows why HARS preserves the target distribution. Steps 1 and 2 implement Theorem 1(ii) with $N = N_\theta + 1$. They preserve the conditional target of the completed draw by the path-reversal argument in Appendix S4. Step 3 proposes the companion block from its unrestricted Gibbs conditional and accepts it if the restrictions hold. This MH step preserves the same conditional target. The latent scales follow the usual BPSS Gibbs update because they do not enter the restrictions. Step 4 preserves the restricted-Haar conditional

distribution of Q by Proposition S2. Proposition S1 then implies that a full HARS iteration preserves the target in Definition 1. HARS avoids the conditional shortcut ruled out by Theorem 1(i). It requires neither the admissible volume $v(\theta)$ nor rejection counts.[3]

Three properties of HARS follow from this block structure. First, the first-block target includes the Jacobian factor $|\det A_0|^T$ and is not Gaussian times an admissibility indicator. HARS updates this block by MH. Second, HARS uses separate updates for these blocks: MH for (A_0, Λ), a Gibbs conditional proposal for A_+, and ESS for the rotation. The sampler exploits this separation. Third, the admissibility weight $\omega_0(z_0; s_0, Q)$ of Theorem 1 applies to the companion block A_+ and thus reweights the marginal target of z_0, which is what makes the completed-draw repair necessary.

HARS treats sign failure as failure of the pair (θ, Q), not of the model-parameter draw θ alone. A draw of model parameters can be inadmissible for the current rotation and admissible for another. Rejecting a proposal at the current rotation therefore does not remove that parameter value from the posterior. The value can be visited later, after the rotation block moves to a rotation under which it is admissible. Parameter values with larger admissible volume are visited more often, which is exactly the weight $v(\theta)$ in Definition 1. HARS thus preserves the heteroskedastic likelihood and uses sign and narrative restrictions only through admissibility.

VI. Illustration

With the normalization on Λ_k imposed by Brunnermeier et al. (2021), both heteroskedastic and homoskedastic models use the same formula for reported impulse responses:

$$\Theta_\ell = \Phi_\ell(B)(A_0^{-1})'Q.$$

Given all the parameter values and the rotation Q, the recovered historical shocks can be calculated as $e_t = Q'\varepsilon_t = Q'\Lambda_{s_t}^{-1/2}A_0'u_t$ for the heteroskedastic model and $e_t = Q'\varepsilon_t = Q'A_0'u_t$ for the homoskedastic model. Let q_j be column j of Q. Then $e_{j,t} = q_j'\varepsilon_t$. $\Theta_{ij,\ell}$ is the reported response of variable i at horizon ℓ to $e_{j,t} = 1$, with the other shocks set to zero. Remember that rows index observables and columns index shocks. Response restrictions are verified for the reported responses and in every regime.

To understand how heteroskedastic and homoskedastic models differ in response probabilities and credible intervals when rotations are applied to satisfy sign and narrative restrictions, we use two analytical closed-form examples with two variables, no lags or deterministic terms, and two regimes. The historical-contribution narrative restriction in Example 2 uses

[3]If posterior-draw repair finds no admissible candidate, the chain stays at its current admissible state. This preserves the target. If the MCMC chain explores the posterior too slowly, we can add a refresh of Q on a fixed schedule. This step holds θ fixed and draws Q from Haar measure restricted to $R(\theta)$. It also preserves the target.

the prevailing-regime impact response and the recovered shocks in the heteroskedastic model and the constant-covariance counterparts in the homoskedastic model.

To keep the examples tractable, we omit the restriction used in the monetary and fiscal applications that the identified shock's absolute historical contribution exceed the sum of the absolute contributions of all other shocks. Even without this additional restriction, the two specifications assign materially different conditional probabilities to unrestricted responses, as shown below.

In each example, $y_t'A_0 = \eta_t'$, $u_t = y_t$, and $\eta_t = A_0'u_t$. The model parameters are fixed; Q remains random. The heteroskedastic model assumes $\eta_t \mid s_t = k \sim \mathcal{N}(0, \Lambda_k)$ independently across dates; the homoskedastic model sets $\Lambda_k = I_2$ in every regime. Before imposing economic restrictions, draw Q from normalized Haar measure on $\mathcal{O}(2)$ by writing

$$Q = \begin{pmatrix} \cos\psi & -\iota\sin\psi \\ \sin\psi & \iota\cos\psi \end{pmatrix}, \qquad \psi \sim \text{Uniform}(-\pi, \pi], \qquad \Pr(\iota = -1) = \Pr(\iota = 1) = \tfrac{1}{2}.$$

Both ψ and ι are random and independent, and ι is a random sign taking only the two values -1 and 1. This construction gives $Q'Q = I_2$, $\det Q = \iota$, and covers all 2×2 orthogonal matrices. The first column is $q_1 = (\cos\psi, \sin\psi)'$. Conditioning on the economic restrictions gives the restricted-Haar distributions used in the following examples; the admissible angle intervals are derived separately for each example.

We set $e_t = (1, 0)'$ for the response experiment. All intervals are pointwise, equal-tailed 90% intervals. The complete model, shock recovery, and every probability and quantile calculation are detailed in Appendices S2–S3.

Example 1 (Sign restrictions only). Let

$$A_0 = \begin{pmatrix} -2/3 & 5/3 \\ 5/3 & -5/3 \end{pmatrix}, \qquad \Lambda_1 = \text{diag}(8/5, 2/5), \qquad \Lambda_2 = \text{diag}(2/5, 8/5).$$

We impose the sign restriction that shock 1 raises y_1 in the reported impulse response and in every regime. With $q_1 = (\cos\psi, \sin\psi)'$, the sign verifications are

$$\cos\psi + \sin\psi > 0, \qquad 2\cos\psi + \sin\psi > 0, \qquad \cos\psi + 2\sin\psi > 0.$$

The last two are the heteroskedastic regime restrictions. Restricted Haar measure makes ψ uniform on $(-\arctan(1/2), \pi/2 + \arctan(1/2))$ under the heteroskedastic model and on $(-\pi/4, 3\pi/4)$ under the homoskedastic model. The unrestricted response of y_2 has the same formula

$$\Theta_{21,0} = \cos\psi + \tfrac{2}{5}\sin\psi$$

in both models.

Analytical results for Example 1

Model	Median	$P(\Theta_{21,0} > 0)$	Equal-tailed 90% interval	Width
Heteroskedastic	0.8737	0.9667	[0.0450, 1.0749]	1.0300
Homoskedastic	0.7616	0.8711	[−0.2642, 1.0737]	1.3379

The heteroskedastic interval is entirely positive; the homoskedastic interval contains zero. Negative y_2 responses remain admissible in both models. The difference is entirely due to verifying the same y_1 sign requirement in every regime for the heteroskedastic model.

Example 2 (The narrative restriction that creates the difference). Let

$$A_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \Lambda_1 = \mathrm{diag}(32/17, 2/17), \qquad \Lambda_2 = \mathrm{diag}(2/17, 32/17).$$

Both models have the same first-shock responses

$$\Theta_{11,0} = \sin\psi, \qquad \Theta_{21,0} = \cos\psi,$$

where the first column of the Haar rotation is $q_1 = (\cos\psi, \sin\psi)'$. The sign restriction is that shock 1 raises y_2 in the reported responses and in every regime. All these verifications reduce to $\cos\psi > 0$, so that ψ is uniform on $(-\pi/2, \pi/2)$ under both models. Without any narrative restriction, the response probabilities are identical.

Suppose the observed residual at date τ is $u_\tau = (-1, 1/4)'$, so that $A_0'u_\tau = (1/4, -1)'$. In both models, the narrative restriction is that the historical impact contribution of shock 1 to y_2 at date τ is positive. For the heteroskedastic model, suppose that τ is in regime 1. Given the response sign verifications, the narrative restriction implies

$$\tfrac{1}{16}\cos\psi - \sin\psi > 0 \quad \text{under the heteroskedastic model,}$$
$$\tfrac{1}{4}\cos\psi - \sin\psi > 0 \quad \text{under the homoskedastic model.}$$

Thus, ψ is uniform on

$$(-\pi/2, \arctan(1/16))$$

under the heteroskedastic model and on

$$(-\pi/2, \arctan(1/4))$$

under the homoskedastic model. The analytical results presented below compare the response before and after imposing the narrative restriction. All intervals are pointwise, equal-tailed 90% intervals.

Analytical results for Example 2

Restrictions and model	Median	$P(\Theta_{11,0} < 0)$	90% interval	Width
Signs only: both models	0.0000	0.5000	[−0.9877, 0.9877]	1.9754
Signs and narrative: H	−0.6847	0.9618	[−0.9967, −0.0192]	0.9774
Signs and narrative: O	−0.6154	0.8651	[−0.9959, 0.1536]	1.1495

The sign restrictions alone give identical response probabilities and intervals in both models. Positive y_1 responses remain admissible in both models. The difference arises when the same narrative restriction is evaluated using the historical shocks recovered under each model. The heteroskedastic interval is entirely negative, whereas the homoskedastic interval contains zero.

Summary and connection to the posterior intervals. Both examples use identical impulse response formulas across models. In Example 1, the heteroskedastic specification verifies the sign restriction in every regime and in impulse responses with normalized shock volatility. In Example 2, the sign restriction gives identical response distributions in both models, but the same narrative restriction selects different rotations because each model recovers a different historical shock at date τ. These differences influence whether the equal-tailed 90% interval of an unrestricted response excludes zero.

With the parameter values used in these examples, the heteroskedastic intervals are narrower. This ordering need not hold generally. In the empirical applications in Section VII, heteroskedastic credible intervals also tend to be narrower than their homoskedastic counterparts, but not necessarily for every response or horizon.

For each vector of model parameters θ with $v(\theta) > 0$, let $K(\theta; \mathcal{A})$ be the conditional probability of a response event $\mathcal{A}$ under Haar measure restricted to the admissible rotations at θ. Under the restricted joint target in Definition 1,

$$\Pr(\Theta_\ell \in \mathcal{A} \mid Y) = \int K(\theta; \mathcal{A})\, \Pi(d\theta \mid Y), \tag{7}$$

where Π is the parameter marginal of this target, *including the admissible-rotation volume weight*. The conditional response probabilities, represented by $K(\theta; \mathcal{A})$, are therefore the building blocks of posterior probabilities. The examples illustrate how these probabilities can differ substantially between the heteroskedastic and homoskedastic specifications and, through (7), affect posterior response probabilities and intervals. The three empirical applications below show such differences.

VII. Empirical Applications

We apply HARS to three SVAR applications. The first is a monetary policy application, in which we identify only the monetary policy shock. This case has a large sign-admissible set. The second is an oil market application. In this application, sign restrictions apply to several shocks and an elasticity bound couples impact responses through a ratio. The third is a fiscal policy application. Sign restrictions are combined with narrative restrictions on two fiscal shocks. In each case, we report the data, model, identifying restrictions, and

impulse responses. Credible intervals for the dynamic impulse responses are computed from HARS draws.[4]

VII.1. **Monetary policy.**

VII.1.1. *Data and setup.* The monetary policy application follows ADRR. They identify a contractionary monetary policy shock with signs and the Volcker narrative restriction. Their monthly U.S. VAR has 12 lags and six variables: real output, the GDP deflator, the commodity price index, total reserves, non-borrowed reserves, and the federal funds rate. The sample is 1965:01–2007:11. We use exactly ADRR's data, specification, and restrictions. We then incorporate regime-dependent shock variances into their model.

We follow the flat-prior tradition of Uhlig (2005), ADRR, and CMT and use

$$p(A_0) \;\propto\; \left|\det(A_0)\right|^{-n},$$

with a flat prior on A_+. The identifying restrictions are imposed only through Q.

We follow BPSS and use four variance regimes with breaks at 1979:10, 1983:01, and 1990:01 (Table 1). BPSS distinguish the Volcker disinflation of 1979–1982 from the S&L crisis period of 1983–1989. ADRR use October 1979 for two narrative restrictions. The monetary policy shock must be positive. Its absolute contribution to the unexpected change in the federal funds rate must also exceed the sum of the absolute contributions of all other shocks.

VII.1.2. *Sign and narrative restrictions.* Table 2 lists the sign restrictions, which hold for six months. A contractionary monetary policy shock raises the federal funds rate and lowers the GDP deflator, commodity prices, and non-borrowed reserves. These restrictions are imposed on one column of Q.

The narrative restriction concerns the date 1979:10. The October 1979 monetary shock must be positive. Its absolute contribution to the unexpected change in the federal funds rate must exceed the sum of the absolute contributions of all other shocks. The Volcker narrative restriction does not impose a decline in output. It does not impose a decline in total reserves. Those responses are outcomes of the estimated SVAR.

VII.1.3. *Policy transmission.* Figure 1 compares the homoskedastic ADRR specification with the heteroskedastic specification under four variance regimes. Blue denotes the homoskedastic specification; red denotes the heteroskedastic specification. Solid lines are posterior medians, and bands are 90% credible intervals. The federal funds rate rises, while

[4]For each application, the impulse-response horizon matches the corresponding study in the existing literature, so that results are compared on the same horizon. Appendix S5 provides the data sources. Section VIII reports the computational tally and the comparison with CMT. Appendix S6 details the sampler settings, hardware, posterior-draw repair steps, and computation time.

the GDP deflator, commodity prices, and non-borrowed reserves fall over the restricted six-month window. Output and total reserves are unrestricted. Their responses show what the model implies beyond the identifying restrictions.

The GDP response is the most illustrative. It is also the central subject in ADRR's monetary policy application. ADRR show that adding their Volcker narrative restrictions to the sign restrictions of Uhlig (2005) makes output fall after about one and a half years at the 68% level. This result is the main economic force of their monetary exercise: narrative information turns an uninformative monetary policy shock into a contractionary policy disturbance with real effects.

The homoskedastic 90% GDP band, however, includes zero at every horizon, indicating that the output contraction is not statistically significant at this conventional level. In contrast, under the heteroskedastic specification, it lies below zero after about two years. At 24 months, the posterior probability of a negative GDP response is 0.97, compared with 0.92 under the homoskedastic specification. At 36, 48, and 60 months, the heteroskedastic probabilities are between 0.97 and 0.98; the homoskedastic probabilities are only 0.90, 0.88, and 0.87 (Table 3, Panel A). The contraction therefore is statistically significant at the 90% level only under the heteroskedastic specification.[5]

Panel B of Table 3 compares responses to a 25-basis-point impact increase in the federal funds rate. At 24 months, for instance, the median GDP response is -0.17% under the heteroskedastic specification and -0.09% under the homoskedastic specification. Their 90% intervals are $[-0.30, -0.02]$ and $[-0.18, 0.02]$. The heteroskedastic interval is wider, yet it excludes zero. At 36, 48, and 60 months, the heteroskedastic specification also gives a larger median contraction and a 90% interval entirely below zero, while the homoskedastic interval contains zero.

The GDP response is the most illustrative because it changes the interpretation of monetary policy. Note that the narrative restriction itself does not constrain output. Under the homoskedastic ADRR specification, the conventional 90% band does not establish an output contraction. Under shock heteroskedasticity, the same narrative information delivers a contraction according to the conventional credibility standard. As a result, the contraction is a credible outcome of the posterior distribution.

The 90% credible bands are informative in two respects. They are more informative statistically, and they change the economic conclusion. Following Sims and Zha (1999), the literature often reports 68% bands when 90% bands are too wide to be useful. That tradition should not substitute for examining the variance and innovation assumptions. A wide band is not just a presentation problem. It can signal that the fitted model can assign uncertainty to the wrong source.

[5]This difference is not a sampler effect. Running HARS with a single variance regime yields impulse responses essentially identical to the homoskedastic ADRR result.

BPSS show that their data strongly favor changing shock variances and fat-tailed innovations. Large innovations occur periodically, especially during episodes such as the Volcker disinflation. A homoskedastic Gaussian SVAR forces tranquil and turbulent periods to share the same shock variances. It uses one variance for each shock to fit both ordinary and extraordinary disturbances. Large historical shocks can then distort inference about dynamic propagation, leaving credible bands uninformative.

The heteroskedastic specification allows large historical disturbances without making all periods equally volatile. The propagation coefficients B are common across regimes, but regime variances enter before rotation and can change the relative impacts of an economic shock. The empirical comparison reflects the innovation distributions and the economic restrictions verified across regimes and at the narrative dates. The GDP result changes the economic interpretation, not just interval width.

Figure 1 gives a median GDP deflator response of about -0.3% after five years under both specifications. The heteroskedastic 90% band is narrower, and its upper band remains farther below zero beyond the restricted six-month window, providing stronger evidence of a persistent price decline after a monetary tightening. For non-borrowed reserves, the homoskedastic upper band crosses zero as soon as the six-month sign restriction ends. The heteroskedastic band remains below zero through month 16. Table 3 confirms this contrast. At months 9, 12, and 15, for instance, the probabilities below zero are 0.97, 0.97, and 0.96 under the heteroskedastic specification, compared with 0.81, 0.80, and 0.81 under the homoskedastic specification (Table 3, Panel A). At 12 months, a 25-basis-point tightening lowers median non-borrowed reserves by 0.39%, with a 90% interval of $[-0.79, -0.04]$. The homoskedastic median is -0.29%, but its interval $[-0.87, 0.27]$ includes zero (Table 3, Panel B). The heteroskedastic specification sustains a credible reserve withdrawal beyond the imposed sign window. The homoskedastic specification does not.

Total reserves are unrestricted. Six months after the shock, the posterior probability that total reserves fall is 0.96 under the heteroskedastic specification and 0.83 under the homoskedastic specification (Table 3, Panel A). For a 25-basis-point tightening, the median responses are -0.22% and -0.24%, with 90% intervals of $[-0.51, -0.01]$ and $[-0.68, 0.17]$. The difference is the probability of a reserve drain, not a larger median decline. Nine and twelve months after the shock, the probabilities of a negative response are 0.94 and 0.93 under the heteroskedastic specification, compared with 0.82 and 0.81 under the homoskedastic specification.

The total-reserves result is also historically meaningful for the 1965:01–2007:11 sample. Before the introduction of interest on reserves in October 2008, the Federal Reserve implemented contractionary policy primarily through open-market operations that drained reserves from the banking system. A rise in the federal funds rate together with a fall in reserves is consistent with a contractionary monetary policy shock under the pre-2008 reserve-operating framework. Belongia and Ireland (2024) formalize this link by identifying

a contractionary policy shock as a joint rise in the federal funds rate and a fall in the supply of bank reserves, and they document that total reserves respond to policy across the pre-2008 regimes. The recursive evidence of Christiano et al. (1996) makes the same point, listing a decline in total reserves among the defining effects of a contractionary monetary policy shock.

The negative reserve responses are consistent with the pre-2008 reserve-operating framework. The heteroskedastic specification gives this reserve drain high posterior probability at six months and generates the decline in non-borrowed reserves beyond the imposed sign horizon. The homoskedastic intervals leave both unrestricted responses uncertain.[6]

Output falls, the price level falls beyond the restricted horizons, and the early reserve responses support a withdrawal of bank reserves. The output result addresses the debate in ADRR and Uhlig (2005). The heteroskedastic price band is narrower, with its upper bound farther below zero. This strengthens the contractionary interpretation of the monetary policy shock. The negative reserve responses reflect the pre-2008 operating framework. These are economic differences between the two specifications, rather than cosmetic changes in credible bands.

In summary, we compare both specifications using 90% credible bands to assess whether the evidence for an output contraction extends beyond the 68% level often used in the literature. Under the same sign and narrative restrictions, the contraction is credible at the 90% level only under the heteroskedastic specification. The narrative restrictions help identify the Volcker shock without imposing an output decline. The heteroskedastic specification strengthens the evidence for its contractionary effect. The price and reserve responses reinforce this interpretation. HARS makes these posterior comparisons feasible.

VII.2. **Oil market.**

VII.2.1. *Data and setup.* KM identify oil market shocks with sign restrictions and an impact price-elasticity bound. We study their monthly trivariate SVAR with oil production growth Δq_t, real global activity y_t, and the real oil price p_t. The SVAR has 24 lags. We use ADRR's extended 1971:01–2015:12 sample and impose KM's sign restrictions and impact price-elasticity bound. We do not impose ADRR's additional narrative restrictions because this comparison uses KM's identifying restrictions. The later computational comparison with CMT in Section VIII uses the same KM restrictions, while the two approaches retain their respective posterior targets. For the heteroskedastic specification, we use regime-dependent shock variances and Student-t innovations as in the monetary policy application.

The prior is the same as in the monetary policy application: flat in A_+ and $p(A_0) \propto |\det(A_0)|^{-n}$. We use three variance regimes. Following Appendix D of CMT, we set the break dates at 1990:01 and 2008:01; Table 4 reports the dates. The sample is therefore

[6]These responses are robust to other regime dates: adding more break dates to the existing ones leaves both the negative GDP response and the negative total-reserves response intact.

divided into three periods: the oil-shock and disinflation era, the Great Moderation, and the financial-crisis/zero-lower-bound period.

VII.2.2. *Sign and elasticity restrictions.* Table 5 provides the details of the identifying restrictions. The sign restrictions are imposed only on impact responses. An oil-supply shock lowers oil production and raises the oil price. An aggregate-demand shock raises production, activity, and the oil price. An oil-specific-demand shock raises production and the oil price but lowers activity. The restrictions are imposed on the impact responses at the BPSS average scale, $(A_0^{-1})'Q$, at which the responses are reported, and in each of the three regimes, $(A_0^{-1})'\Lambda_k^{1/2}Q$, with a common orientation of each shock. Because $\Lambda_k^{1/2}$ enters before the rotation, the regime variances affect both sign admissibility and the elasticity ratio discussed below. The requirement therefore binds: a single rotation must satisfy the KM restrictions in every regime.

The elasticity bound of 0.0258 applies to the two demand shocks. For each demand shock, it limits the ratio of the impact response of oil production growth to the impact response of the real oil price. The bound restricts how much oil production growth can move within the month when oil demand shifts. It does not apply to the oil-supply shock. For that shock, the sign restriction keeps the oil-price response positive, while the elasticity bound leaves its magnitude unrestricted. Thus, the bound rules out rotations that imply an implausibly elastic short-run oil-supply curve.[7] Because the bound is a ratio, it links two impact responses for the same demand shock: the response of oil production and the response of the real oil price. This sharply reduces the admissible set and makes traditional accept–reject exploration costly. HARS avoids that cost.

VII.2.3. *Economic implications.* Figure 2 compares the homoskedastic specification in blue with the heteroskedastic specification in red. Following ADRR, we use an 18-month horizon for both specifications. Solid lines are posterior medians, and bands are 90% credible intervals.

Both specifications yield the imposed impact signs. An oil-supply shock lowers oil production and raises the real oil price. An aggregate-demand shock raises production, real activity, and the real oil price. An oil-specific-demand shock raises production and the real oil price but lowers real activity on impact. These impact responses are dictated by the identifying restrictions. The economic content lies in the unrestricted horizons and in the relative magnitudes across shocks. KM's central claims are about *which* shock contributes most to movements in the real price of oil and by *how much* relative to the others. The homoskedastic and heteroskedastic specifications differ on exactly these dimensions and in the direction KM's economic argument requires.

[7]Recent evidence supports a small short-run oil-supply elasticity. Kilian and Patel (2026) use the 2026 Iran War as a natural experiment and find that one-month and one-quarter U.S. oil supply elasticities are effectively zero. Their findings support tight elasticity bounds even in the shale-oil era.

Precision. The heteroskedastic credible bands are generally sharper than the homoskedastic ones, the same finding as in the monetary policy application. We therefore do not repeat that discussion here. KM present 68% credible bands. We keep the conventional 90% standard throughout, and the heteroskedastic specification makes the responses informative at that standard.[8] The remaining results show that this sharpening is not uniform. It matters most where KM's economic argument and the homoskedastic estimates disagree.

Real oil price. The homoskedastic median real-oil-price response to a one-standard-deviation oil-supply shock peaks at about 2.2% in month 5, with a 90% interval of $[0.6, 3.8]$%. At the same horizon, the heteroskedastic median is close to zero, with an interval of $[-0.5, 0.8]$%. The upper endpoint of the heteroskedastic 90% interval remains below 1% at every horizon beyond impact in Figure 2. The heteroskedastic posterior therefore places little probability on a price increase as large as the homoskedastic median at five months. This finding supports KM's argument that the oil-price effects of supply shocks are limited.

Under the homoskedastic specification, the 90% oil-price interval remains above zero throughout the horizon for an aggregate-demand shock. For an oil-supply shock, it is above zero at months 2–6, in addition to impact. For an oil-specific-demand shock, it remains above zero only through month 3. The evidence for a positive oil-specific-demand price effect therefore ends earlier than for either of the other shocks. Under the heteroskedastic specification, this comparison with oil supply is reversed: the oil-specific-demand interval remains above zero from impact through month 7, while the oil-supply interval contains zero at every horizon beyond impact. Shock heteroskedasticity thus provides evidence for a persistent oil-price increase after an oil-specific-demand shock while limiting the supply-price effect.

The two demand shocks raise the real oil price under both specifications. Forecast-error variance decompositions establish their dominant contribution to oil-price variation.[9] At the six-month forecast horizon, for example, the median oil-supply share is 3% under the homoskedastic specification and below 1% under the heteroskedastic specification. The upper endpoints of their 90% intervals are 9% and 1%. The heteroskedastic specification also places a tighter upper bound on the oil-supply contribution.

KM restrict the oil-price response only on impact. The positive oil-specific-demand response through month 7 under the heteroskedastic specification therefore extends beyond the imposed restriction. At the six-month forecast horizon, the median oil-specific-demand

[8] With a single variance regime and the final repair rule, HARS yields impulse responses essentially identical to the homoskedastic KM result.

[9] Variance shares are computed draw by draw. The comparison across specifications sets $\Lambda = I_n$ in the heteroskedastic model, as in Figure 2. The calculations by regime use that regime's shock variances throughout the forecast horizon.

share is 48% under the heteroskedastic specification, compared with 35% under the homoskedastic specification. The aggregate-demand medians are 52% and 62%.[10] Overall, the heteroskedastic specification assigns a smaller median share to aggregate demand and a larger one to oil-specific demand than the homoskedastic specification (Table 6).

Across the forecast horizons of months 1, 6, 12, and 18 and under the heteroskedastic specification, oil-specific demand has the largest median share of oil-price forecast-error variance in 1990–2007, ranging from 60% to 82% (Table 7). In 2008–2015, aggregate demand has the largest median share, ranging from 58% to 80%. For 1990–2007, oil-specific demand contributes more to oil-price forecast-error variance than aggregate demand at all four reported horizons in the median estimates. The homoskedastic specification instead assigns aggregate demand the largest median share at all four horizons. Shock heteroskedasticity thus reveals a leading role for oil-specific demand in the 1990–2007 median decomposition, a role absent from the homoskedastic estimates. It also generates a positive oil-price response to this shock at the 90% level through month 7, compared with month 3 under the homoskedastic specification.

Economic contraction. An oil-specific demand shock captures precautionary, expectations-driven demand for oil. It raises the real price of oil within the month and, under KM's economic interpretation, should depress global real activity in the short run. KM do not obtain this contraction with a credible band. Their preferred point estimate implies the contraction—the corresponding impact coefficient in their final model is clearly negative—but except at impact, where the sign is imposed, the credible bands in their Figure 5 are too wide for the rest of the horizon even at the 68% level. In our homoskedastic specification, the early 90% band excludes zero at impact and at month 1, but includes zero at months 2–9. Its upper bound is approximately zero at month 10.

The heteroskedastic 90% band remains below zero through month 3. At months 2 and 3, the contraction remains credible at the conventional 90% level, while the homoskedastic band contains zero. The upper bound is only slightly above zero at months 4 and 5 and falls below zero again at month 6. An oil-specific-demand shock thus raises the real oil price and depresses real activity beyond the imposed impact signs.[11] That is the heart of KM's economic argument.

Summary. Shock heteroskedasticity reveals a leading role for oil-specific demand in the 1990–2007 median variance decomposition, a role absent from the homoskedastic estimates.

[10] At the six-month forecast horizon, the heteroskedastic median oil-specific-demand share is 77% in 1990–2007, compared with 22% for aggregate demand. In 2008–2015, the median aggregate-demand share is 65% and the oil-specific-demand share is 35%. The homoskedastic specification instead assigns the same median shares in every period: 62% to aggregate demand and 35% to oil-specific demand.

[11] Both specifications also have 90% real-activity bands below zero at months 11–15. ADRR add narrative restrictions and find that the contraction beyond impact becomes significant at the 68% level only after about 18 months.

Its positive oil-price effect remains credible at the 90% level through month 7, compared with month 3 under the homoskedastic specification. The oil-supply price effect is small and not credibly positive beyond impact under the heteroskedastic specification. Both specifications find a later contraction in real activity. The heteroskedastic specification also produces the early contraction through month 3, when the oil-price band is above zero and the real-activity band is below zero. These simultaneous and unrestricted responses support KM's economic interpretation: an oil-specific-demand shock raises oil prices and contracts real activity in the early months.[12]

VII.3. **Fiscal shocks.** Unlike the monetary policy and oil market applications, which closely replicate the identifying restrictions of ADRR and KM, the fiscal policy application combines sign and narrative restrictions from several leading approaches in the fiscal SVAR literature. We identify a net-tax shock and a government-spending shock and study how shock heteroskedasticity changes inference about their economic effects.

The fiscal application is the most demanding of the three. Computationally, it combines the richest set of identifying restrictions: multi-shock sign restrictions, narrative restrictions, and shock heteroskedasticity. Empirically, postwar U.S. fiscal policy and macroeconomic fluctuations differ markedly across episodes, making regime-dependent shock variances a natural part of the model.

VII.3.1. *Data.* We follow Caldara and Kamps (2017), who implement the sign-restriction identification of Mountford and Uhlig (2009) in a compact five-variable quarterly U.S. SVAR with four lags. The system contains the three fiscal variables in Blanchard and Perotti (2002)—net taxes, government spending, and real GDP—augmented with the Treasury bill rate and the GDP deflator as the price level. The T-bill rate has a longer sample than the federal funds rate, allowing us to start in the late 1940s and retain the early fiscal episodes.

Net taxes follow Blanchard and Perotti (2002) and Mountford and Uhlig (2009), who treat transfers as negative taxes. Thus,

$$\mathrm{NT}_t = \mathrm{REC}_t - \mathrm{TR}_t - \mathrm{INT}_t,$$

where NT_t is net taxes, REC_t is total government current receipts, TR_t is government current transfer payments, and INT_t is government interest payments. The three components on the right-hand side are consolidated federal, state, and local government series. Government spending is nominal government consumption expenditures and gross investment for all levels of government. This purchases measure excludes transfers. Net taxes, government spending, and GDP are divided by NIPA midperiod population and deflated by the GDP deflator.

[12]Our results are robust to the regime specification. Adding more break dates to the original CMT dates leaves intact the muted price response to an oil-supply shock, the relative importance of the three oil shocks, and the short-run economic contraction.

The log transformation for net taxes requires $\mathrm{NT}_t > 0$. This condition holds throughout 1947Q1–2007Q4.

In this SVAR, net taxes, government spending, and real GDP enter as real per-capita logs; the Treasury bill rate enters as a decimal level; and the GDP deflator enters as a log level. The sample period is 1947Q1–2007Q4, which contains all six narrative episodes from Romer and Romer (2010) and Ramey (2011). Neither study provides narrative episodes beyond 2007.

VII.3.2. *Identification.* We identify two fiscal shocks: a net-tax shock and a government spending shock. Unlike Mountford and Uhlig (2009), we identify the effects of these two shocks simultaneously in the five-variable system.[13]

Sign restrictions. Table 8 summarizes the sign restrictions. All sign restrictions are imposed on impact and over the following three quarters—the four-quarter window of Mountford and Uhlig (2009) and Caldara and Kamps (2017). In the spending specification, the government spending shock raises both government spending and GDP, with net taxes left unrestricted.

In the tax specification, the net-tax shock raises net taxes and contracts GDP, with government spending left unrestricted. The spending shock raises GDP, following Jørgensen and Ravn (2022), and the net-tax shock lowers GDP. The negative output response imposed on the net-tax shock is well supported across alternative identifications in the literature. For the tax component, Blanchard and Perotti (2002) estimate that a positive tax shock lowers output within the quarter, and Romer and Romer (2010) find that an exogenous tax increase of one percent of GDP reduces real GDP by close to three percent. For the transfer component, Arias et al. (2026a) emphasize that fiscal transfer shocks sustain real output. Thus, a cut in transfers, which is a positive net-tax innovation, is contractionary. For both fiscal shocks, we leave the Treasury bill rate and the GDP deflator unrestricted.

Narrative restrictions. Narrative sign restrictions are reported in Panel (b) of Table 8. At each specified date, the historical fiscal shock has the restricted sign. Its absolute contribution to the unexpected change in its own policy variable exceeds the sum of the absolute contributions of the other shocks. We now describe the timing of these narrative sign restrictions.

The tax and spending restrictions are based on different narrative sources. For taxes, Romer and Romer (2010) identify so many exogenous tax changes that we select one tax cut and one tax increase. The tax cut is the Revenue Act of 1964, the Kennedy–Johnson reform motivated by long-run economic growth, dated 1964Q2 and assigned a negative shock. The tax increase is the Omnibus Budget Reconciliation Act of 1993 (OBRA-93), dated 1993Q4

[13] Mountford and Uhlig (2009) identify business cycle and monetary policy shocks to help isolate fiscal policy shocks. These auxiliary business cycle and monetary policy shocks are not needed in our application because the sign and narrative restrictions specified below directly identify the two fiscal policy shocks.

and assigned a positive shock. Romer and Romer (2010) classify it as motivated by deficit reduction. Together, these two episodes identify the net-tax shock in both directions; they are the two clearest categories: one for a tax cut and one for a tax increase.[14]

For the spending shock, we use all four military build-ups documented by Ramey (2011). Because the identified shock is an innovation to observed government spending, we place each restriction in the quarter where spending surges in our data, within four quarters of the military build-up announcement. The four episodes are the Korean War in 1951Q1, the Vietnam War in 1965Q3, the Carter–Reagan build-up in 1980Q1, and the post-9/11 build-up in 2001Q4. The identified innovation is relative to the VAR information set. Dating the restriction at implementation does not necessarily mean that private agents received no earlier news. The historical spending shock is restricted to be positive in all four episodes.

Variance regimes and priors. The heteroskedastic model has three variance regimes, with breaks at 1971Q2 and 1997Q1. These are the fiscal regimes used in Appendix D of CMT. Their fiscal model is a heteroskedastic proxy-SVAR, not our sign-and-narrative design, but the volatility regimes are part of their fiscal sample. Table 9 reports the dates. The prior is the same as in the monetary policy and oil market applications: flat in A_+ and $p(A_0) \propto |\det(A_0)|^{-n}$.

VII.3.3. *Policy insights.* Figures 3 and 4 compare responses to the same initial fiscal changes under the two specifications. The net-tax shock raises net taxes by 1% on impact, and the spending shock raises government spending by 1% on impact. We normalize each posterior draw before computing the medians and pointwise equal-tailed 90% intervals. Blue denotes the homoskedastic specification and red the heteroskedastic specification. Quarter 0 is impact. The sign restrictions apply to quarters 0–3; the price and interest-rate responses are unrestricted.

Net-tax shock, output, and prices. The initial net-tax increase is identical across specifications, but its persistence differs. The heteroskedastic 90% net-tax band remains above zero through quarter 6, compared with quarter 3 under the homoskedastic specification (Figure 3). At quarter 6, the median net-tax increase is about 0.42% under the

[14]Romer and Romer's narrative appendix dates the 1964 tax cut to 1964Q2. For OBRA-93, it records a positive tax-liability change in 1993Q4 excluding retroactivity and a negative change including retroactivity. The positive restriction at 1993Q4 refers to the former implementation measure. The sign and dominance of the aggregate net-tax innovation remain identifying assumptions; the enactment date does not establish that the innovation was unanticipated. We use Romer and Romer (2010) to date the tax actions, not to compare impulse responses, because our fiscal shock is a net-tax innovation rather than a legislated permanent tax shock. Net taxes are receipts less transfers and interest payments in our data. An increase in taxes or a cut in transfers raises net taxes and reduces disposable income, holding other flows fixed. This channel is consistent with the fiscal-transmission discussion in Ramey (2016). See also Perotti (2005, fn. 4).

heteroskedastic specification and close to zero (0.03%) under the homoskedastic specification. The government-spending and Treasury-bill-rate bands contain zero throughout the horizon under both specifications.

Despite the difference in net-tax persistence, both specifications produce a persistent output contraction. The heteroskedastic GDP median troughs at about -0.20% around quarter 12, compared with about -0.19 around quarter 9 under the homoskedastic specification. Both 90% GDP bands remain below zero through quarter 19, well beyond the restricted first year. The heteroskedastic interval can be wider. At quarter 12, for example, the heteroskedastic median and interval are -0.20% and $[-0.35\%, -0.08\%]$, compared with -0.17% and $[-0.30\%, -0.06\%]$ under the homoskedastic specification. This persistent contraction is consistent with the tax-transmission evidence in Ramey (2016) and Mertens and Ravn (2013).

At the end of the forecast horizon (quarter 19), the heteroskedastic GDP-deflator median is about -0.26%, with a 90% interval of $[-0.51, -0.02]\%$. The homoskedastic median is -0.18%, with an interval of $[-0.38, 0.02]\%$. The heteroskedastic interval is wider, yet it lies entirely below zero. The heteroskedastic price decline is credible at the 90% level at quarters 18 and 19, while the homoskedastic band contains zero throughout. The distinction comes from where posterior probability lies, as illustrated in Section VI.

Government spending shock, GDP, net taxes, and prices. A spending shock raises government spending by 1% on impact under both specifications (Figure 4). The median spending increase peaks around quarter 3 at about 1.43% under the heteroskedastic specification and 1.56% under the homoskedastic specification. By quarter 19, both medians are near 0.5%. Both 90% spending bands remain above zero throughout; the spending increase is persistent under either specification. Meanwhile, the Treasury-bill-rate band contains zero throughout under both specifications.

Despite these similar spending paths, the output responses differ. The heteroskedastic GDP median peaks near 0.24% at quarter 2 and declines to about 0.08% at quarter 19. The homoskedastic median reaches about 0.33% around quarter 10 and remains near 0.26% at quarter 19. The heteroskedastic GDP band stays above zero only through quarter 5, compared with quarter 17 under the homoskedastic specification. Only quarters 4 and 5 extend beyond the four-quarter sign restriction under the heteroskedastic specification. In contrast to the homoskedastic specification, a persistent spending increase does not produce a persistent output expansion at the 90% level.

The heteroskedastic GDP median moves toward zero much faster than the homoskedastic median. At quarter 12, for example, the heteroskedastic median is 0.13%, with a 90% interval of $[-0.13, 0.40]\%$. The homoskedastic median is 0.31%, with an interval of $[0.05, 0.61]\%$. The widths are similar, but the heteroskedastic median is much closer to zero, and its interval extends well on both sides of zero. The difference between the two specifications concerns

where posterior probability lies, not interval width alone. Section VI illustrates how changes in that location can alter inferences.

The heteroskedastic net-tax band contains zero at every horizon. Under the homoskedastic specification, it remains above zero from quarter 3 through quarter 19. At quarter 12, for instance, the median net-tax response is about 0.44% under the heteroskedastic specification and 1.04% under the homoskedastic specification. A persistent net-tax increase after a spending shock is therefore not robust to the volatility specification.

The homoskedastic GDP-deflator band lies below zero at quarters 6–10 despite the significant output expansion. This is the fiscal price puzzle emphasized by Jørgensen and Ravn (2022). The heteroskedastic GDP-deflator band contains zero throughout. For the same initial spending increase, the credible price decline occurs only under the homoskedastic specification.

In summary, under the heteroskedastic specification, a positive government spending shock produces no fiscal price puzzle at the 90% credible level. There is no credible net-tax increase at any horizon in Figure 4. The GDP expansion remains credible only through quarter 5, two quarters beyond the four-quarter restriction window. The homoskedastic specification instead shows a net-tax increase from quarter 3 through quarter 19, accompanied by a GDP expansion through quarter 17 and a price decline at quarters 6–10 at the 90% credible level—consistent with a fiscal price puzzle.

Posterior-draw repair and fiscal inference. This fiscal application demonstrates that posterior-draw repair is essential in HARS. HARS delivers admissible draws, while our implementation of CMT's Algorithm 2 produces none. Without repair, the HARS chains remain trapped near their starting regions and produce materially different fiscal responses. With repair, chains from different starting points reach the same posterior region and agree on the responses. Table 10 gives the share of stale-Q events resolved by posterior-draw repair.

When a completed candidate fails the restrictions, posterior-draw repair holds the current rotation and latent variables fixed, continues the completed MH draws from the last failed candidate, and retains the first admissible completed draw. Theorem 1 establishes invariance of the restricted target. The repair budget changes the transition kernel while preserving that target.

Without repair, chains started in different regions produce different response intervals over the runs examined. A chain without repair, restarted from the final state of a repaired chain, reproduces the repaired results while remaining in that region. With repair, four chains at $N_\theta = 100$ and four at $N_\theta = 200$, started from the two previously explored regions and two dispersed points, reach the same posterior region and agree on the fiscal responses. Increasing the budget from 100 to 200 does not materially change the MCMC results. Therefore, we use $N_\theta = 100$ in the baseline.

Our implementation of CMT's Algorithm 2 accepts no draws among 200,000 candidates in the fiscal application and therefore supplies no posterior bands for this comparison. CMT and HARS target different posteriors, as discussed above. The candidate count documents computational difficulty in this implementation; it does not establish that CMT's population identified set is empty.

VIII. Computational Efficiency of HARS

In this section, we compare the computational cost of HARS with our implementation of CMT's Algorithm 2. More details are in Appendix S6. The comparisons use the same data, lag specifications, variance regimes, and economic restrictions, but the two approaches retain their respective posterior targets. In both applications, CMT's credible bands are narrower than the HARS bands for most responses. The monetary policy comparison differs in the strength of the price-level response beyond the imposed sign horizon (left panel of Figure 5). The oil comparison differs in the response of economic activity to an oil-specific demand shock (right panel). HARS uses Student-t shocks, following BPSS; CMT's Algorithm 2 assumes Gaussian shocks. These response contrasts therefore reflect different innovation distributions and posterior targets, rather than the sampling algorithms alone.[15]

Under Definition 1, sign and narrative restrictions act on the model parameters only through the admissible volume $v(\theta)$ of their rotations, and no parameter value with positive admissible volume is excluded. CMT's Algorithm 2 instead draws model parameters directly, using the row-by-row sampler of Waggoner and Zha (2003) for the contemporaneous block, and accepts draws only when their implied responses and historical shocks satisfy the imposed restrictions. This restriction is intentional: their posterior relies on identifying information from heteroskedasticity on top of sign and narrative restrictions. This approach has distinctive conceptual as well as computational implications. Computationally, low acceptance rates can make this approach costly or even infeasible in some applications. Conceptually, CMT discuss label switching and the possibility of an empty identified set in their Section 2.5. The reported finite-sample acceptance rates alone do not establish whether the population identified set is unique, multiple, or empty.

Table S4 compares the computing time needed to obtain 1,000 effective impulse response draws. With posterior-draw repair enabled, HARS takes 0.097 hours in the monetary application, 0.034 in oil, and 0.035 in fiscal. CMT's Algorithm 2 takes 6.60 hours in the monetary application and 1.095 in oil. The monetary CMT timing comes from a short 500-draw run and indicates the order of magnitude. In our fiscal implementation, CMT's Algorithm 2 accepts no draws among 200,000 candidates and therefore supplies no posterior bands. The candidate count documents computational difficulty in this implementation; it does not establish that CMT's population identified set is empty.

[15]The complete set of impulse responses is reported in Appendix S6.1.

Table 10 gives the effect of posterior-draw repair across the three applications. A stale-Q event occurs when a model-parameter proposal makes the current rotation inadmissible. Holding Q and the latent variables fixed, repair continues the completed draws of (A_0, Λ, A_+) from the last failed candidate and retains the first admissible one (Theorem 1(ii)). If no admissible candidate is found within the repair budget, the chain retains its current draw. The table compares repair enabled ($N_\theta = 100$) and disabled ($N_\theta = 0$), holding the model, data, restrictions, variance regimes, and remaining sampler settings fixed.

With $N_\theta = 100$, repair resolves 98.0% of stale-Q events in the monetary application, 69.3% in oil, and 91.7% in fiscal. Disabling repair lowers the estimated time per 1,000 effective draws to 0.049, 0.006, and 0.008 hours. These lower times do not establish adequate posterior exploration. A chain trapped in one region can yield misleading economic and statistical inference.

The fiscal restart tests in Section VII.3.3 show why repair is more than a timing improvement. Without repair, chains remain dependent on their starting region and give materially different fiscal responses. With repair, chains from different starts agree at $N_\theta = 100$ and $N_\theta = 200$. The repair budget affects exploration, not the posterior target. An unsuccessful repair retains the current admissible draw, as required by Theorem 1(ii).

IX. Conclusion

HARS uses sign and narrative restrictions to select economically meaningful rotations while leaving the likelihood unchanged. That principle has grounded homoskedastic SVARs in the literature. It is, however, more difficult to maintain under shock heteroskedasticity. Once relative shock variances differ across regimes, rotating the contemporaneous and lagged coefficients while holding the regime variances fixed tilts the heteroskedastic likelihood. It changes the fitted model.

HARS restores the principle. It absorbs regime volatility before rotation, rotates the absorbed block, and leaves the heteroskedastic likelihood unchanged. The BPSS heteroskedastic specification remains the likelihood engine. It provides a computationally efficient way to handle heavy tailed fluctuations in macroeconomic time series. The posterior draw carries the likelihood information. The rotation pins down the economic interpretation. HARS builds on standard MCMC posterior simulators, including the BPSS sampler and the Gibbs sampler, with parameter updates adapted to the restricted joint target. It alternates these updates with restricted-Haar updates of the rotation and repairs a parameter proposal on the completed draw when the current rotation becomes inadmissible. The restrictions enter only through admissibility, never through the likelihood. This separation is the central principle of the paper. It gives sign and narrative restrictions a coherent role in heteroskedastic SVARs and turns sign admissibility into a well-defined sampling problem.

Across the three empirical applications, the heteroskedastic specification alters economic inference under the same identifying restrictions. The differences concern posterior location

and persistence, not interval width alone. In the fiscal application, heteroskedastic intervals can be wider or have similar widths at different locations. The question is which economic responses receive high posterior probability.

In the monetary policy application, the Volcker contraction becomes credible at the conventional 90% level. In the oil market application, the heteroskedastic specification supports a sustained oil-price increase and an early contraction in real activity after an oil-specific demand shock beyond the impact restrictions. The homoskedastic specification assigns aggregate demand the largest median share in every period. The heteroskedastic specification reveals oil-specific demand's leading role in 1990–2007: its median share of real-oil-price forecast-error variance ranges from 60% to 82%, a majority at all four horizons.

In the fiscal application, net-tax increases produce output contractions credible through quarter 19 under both specifications. After a spending increase, the heteroskedastic GDP expansion remains credible only through quarter 5. There is no fiscal price puzzle at the 90% level, and no net-tax increase is credible at any horizon. The homoskedastic specification instead shows a net-tax increase from quarter 3 through quarter 19, accompanied by a GDP expansion through quarter 17 and a price decline at quarters 6–10 at the 90% credible level—a fiscal price puzzle.

HARS makes these comparisons practical without changing the heteroskedastic likelihood or weakening the economic restrictions. Its general MCMC construction, explicit posterior target, and posterior-draw repair provide the methodological foundation. We use established applications, the BPSS variance specification, and fixed regime dates for transparent comparisons. The construction provides a basis for future extensions of HARS to applications with estimated Markov-switching regimes.

TABLE 1. Dates for variance regimes (1965:01–2007:11), 4-regime specification

Regime	Start	End	Description
1	1965:01	1979:09	Pre-Volcker period.
2	1979:10	1982:12	Volcker disinflation.
3	1983:01	1989:12	Post-Volcker transition.
4	1990:01	2007:11	Great Moderation / pre-crisis period.

Notes: The three break dates follow BPSS (Table 2). The January 1983 break separates the Volcker disinflation from the post-Volcker transition. End dates are inclusive. The last regime ends in 2007:11 because the estimation sample ends in 2007:11.

TABLE 2. Sign and narrative restrictions for monetary policy shock

	GDP Deflator	Commodity Prices	Non-borrowed Reserves	Federal Funds Rate
Monetary policy shock	− (6m)	− (6m)	− (6m)	+ (6m)
Narrative: the shock is positive in October 1979 (Volcker disinflation). Its absolute contribution to the unexpected change in the federal funds rate exceeds the sum of the absolute contributions of all other shocks.				

TABLE 3. Responses to a contractionary monetary policy shock

	Months after the shock				
Panel A. Posterior probability that the response is negative					
GDP	12	24	36	48	60
Homo	0.61	0.92	0.90	0.88	0.87
Hetero	0.68	0.97	0.98	0.97	0.97
Total reserves	0	3	6	9	12
Homo	0.65	0.75	0.83	0.82	0.81
Hetero	0.61	0.83	0.96	0.94	0.93
Non-borrowed reserves	6	9	12	15	18
Homo	0.89	0.81	0.80	0.81	0.77
Hetero	1.00	0.97	0.97	0.96	0.93
Panel B. Median and 90% band, per 25bp impact on the funds rate					
GDP	12	24	36	48	60
Homo	−0.02	−0.09	−0.10	−0.11	−0.13
	[−0.15, 0.09]	[−0.18, 0.02]	[−0.22, 0.03]	[−0.27, 0.05]	[−0.32, 0.07]
Hetero	−0.04	−0.17	−0.21	−0.22	−0.22
	[−0.20, 0.11]	[−0.30, −0.02]	[−0.36, −0.05]	[−0.38, −0.04]	[−0.40, −0.03]
Total reserves	0	3	6	9	12
Homo	−0.13	−0.15	−0.24	−0.26	−0.27
	[−0.66, 0.40]	[−0.51, 0.20]	[−0.68, 0.17]	[−0.73, 0.19]	[−0.78, 0.22]
Hetero	−0.03	−0.10	−0.22	−0.25	−0.31
	[−0.20, 0.10]	[−0.33, 0.06]	[−0.51, −0.01]	[−0.59, 0.02]	[−0.71, 0.03]
Non-borrowed reserves	6	9	12	15	18
Homo	−0.27	−0.27	−0.29	−0.30	−0.27
	[−0.72, 0.08]	[−0.80, 0.23]	[−0.87, 0.27]	[−0.87, 0.25]	[−0.86, 0.31]
Hetero	−0.37	−0.32	−0.39	−0.42	−0.40
	[−0.66, −0.15]	[−0.66, −0.04]	[−0.79, −0.04]	[−0.87, −0.02]	[−0.91, 0.05]

Notes: Homo is the homoskedastic model (ADRR); Hetero is the heteroskedastic model with four variance regimes. Panel A reports pointwise posterior frequencies over the retained draws of (θ, Q). They are invariant to shock-size normalization because each draw is scaled by its own positive impact on the federal funds rate. A probability of at least 0.95 means the 90% band lies entirely below zero. In Panel B, each cell reports the posterior median in percent, with the 90% band in brackets below it. All reported responses are unrestricted; non-borrowed reserves are sign-restricted only for months 0–5, which are not shown.

TABLE 4. Dates for variance regimes (1971:01–2015:12), 3-regime specification

Regime	Start	End	Description
1	1971:01	1989:12	Oil shocks, stagflation, and post-disinflation transition.
2	1990:01	2007:12	Great Moderation.
3	2008:01	2015:12	Financial crisis and zero-lower-bound period.

Notes: Break dates follow CMT (App. D), who fix the predetermined regimes to two of the BPSS variance-regime break dates. End dates are inclusive.

TABLE 5. Sign and elasticity restrictions for the oil market SVAR

	Oil supply shock	**Aggregate demand shock**	**Oil-specific demand shock**
Δq_t (Oil production)	$-$	$+$	$+$
y_t (Global activity)	$-$	$+$	$-$
p_t (Real oil price)	$+$	$+$	$+$
Elasticity bound	—	$\Theta_{12,0}/\Theta_{32,0} \leq 0.0258$	$\Theta_{13,0}/\Theta_{33,0} \leq 0.0258$

Notes: The sign restrictions are imposed only on impact responses. The elasticity bound limits the ratio of the impact response of oil production growth to the impact response of the real oil price for each demand shock. Rows 1 and 3 index oil production growth and the real oil price; columns 2 and 3 index the two demand shocks. The same bounds are imposed on each regime impact matrix $\Theta_{k,0}(Q)$.

TABLE 6. Share of real-oil-price forecast-error variance due to each shock

	Horizon (months)			
	1	6	12	18
Oil supply				
Homoskedastic	0.01	0.03	0.04	0.04
	[0.00, 0.03]	[0.00, 0.09]	[0.01, 0.12]	[0.01, 0.12]
Heteroskedastic	0.00	0.00	0.00	0.01
	[0.00, 0.01]	[0.00, 0.01]	[0.00, 0.02]	[0.00, 0.04]
Aggregate demand				
Homoskedastic	0.50	0.62	0.69	0.74
	[0.07, 0.94]	[0.15, 0.95]	[0.23, 0.94]	[0.32, 0.92]
Heteroskedastic	0.45	0.52	0.62	0.70
	[0.09, 0.90]	[0.13, 0.94]	[0.21, 0.95]	[0.28, 0.94]
Oil-specific demand				
Homoskedastic	0.49	0.35	0.26	0.21
	[0.05, 0.92]	[0.02, 0.81]	[0.03, 0.72]	[0.05, 0.64]
Heteroskedastic	0.55	0.48	0.37	0.28
	[0.09, 0.91]	[0.06, 0.87]	[0.05, 0.79]	[0.05, 0.70]

Notes: Entries are posterior medians, with 90% bands in brackets below them, of the share of the k-month-ahead forecast-error variance of the real oil price due to each shock. Shares are computed draw by draw. For the heteroskedastic model, every regime variance is set to its average, $\Lambda = I$. Medians need not sum to one across shocks.

TABLE 7. Share of real-oil-price forecast-error variance due to each shock, by variance regime

	Horizon (months)			
	1	6	12	18
Oil supply				
Homoskedastic	0.01	0.03	0.04	0.04
	[0.00, 0.03]	[0.00, 0.09]	[0.01, 0.12]	[0.01, 0.12]
Heteroskedastic, 1973–1989	0.01	0.01	0.02	0.04
	[0.00, 0.03]	[0.00, 0.04]	[0.00, 0.08]	[0.01, 0.14]
Heteroskedastic, 1990–2007	0.00	0.00	0.00	0.01
	[0.00, 0.00]	[0.00, 0.01]	[0.00, 0.02]	[0.00, 0.04]
Heteroskedastic, 2008–2015	0.00	0.00	0.00	0.00
	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.01]
Aggregate demand				
Homoskedastic	0.50	0.62	0.69	0.74
	[0.07, 0.94]	[0.15, 0.95]	[0.23, 0.94]	[0.32, 0.92]
Heteroskedastic, 1973–1989	0.55	0.62	0.70	0.75
	[0.12, 0.93]	[0.18, 0.96]	[0.28, 0.95]	[0.36, 0.94]
Heteroskedastic, 1990–2007	0.18	0.22	0.30	0.39
	[0.02, 0.72]	[0.04, 0.81]	[0.06, 0.84]	[0.09, 0.82]
Heteroskedastic, 2008–2015	0.58	0.65	0.74	0.80
	[0.13, 0.94]	[0.19, 0.97]	[0.30, 0.97]	[0.40, 0.97]
Oil-specific demand				
Homoskedastic	0.49	0.35	0.26	0.21
	[0.05, 0.92]	[0.02, 0.81]	[0.03, 0.72]	[0.05, 0.64]
Heteroskedastic, 1973–1989	0.44	0.37	0.27	0.19
	[0.06, 0.87]	[0.04, 0.81]	[0.03, 0.69]	[0.03, 0.58]
Heteroskedastic, 1990–2007	0.82	0.77	0.69	0.60
	[0.28, 0.98]	[0.18, 0.96]	[0.15, 0.93]	[0.16, 0.90]
Heteroskedastic, 2008–2015	0.42	0.35	0.26	0.19
	[0.06, 0.87]	[0.03, 0.81]	[0.03, 0.70]	[0.03, 0.60]

Notes: Entries are posterior medians, with 90% bands in brackets below them, of the share of the k-month-ahead forecast-error variance of the real oil price due to each shock. Shares are computed draw by draw. For the heteroskedastic model, each row sets the shock variances to those of the stated regime. Medians need not sum to one across shocks.

TABLE 8. Sign and narrative restrictions for the fiscal SVAR

Panel (a): Sign restrictions

	Net-tax	**Gov. spending**
Net Taxes	+ (4q)	
Gov. Spending		+ (4q)
GDP	− (4q)	+ (4q)
T-bill		
GDP Defl.		

Panel (b): Narrative restrictions

Shock	Quarter	Episode	Realized sign
Net-tax	1964Q2	Revenue Act of 1964 (Kennedy–Johnson tax cut)	−
	1993Q4	OBRA-93 (Clinton tax increase)	+
Gov. spending	1951Q1	Korean War build-up	+
	1965Q3	Vietnam War build-up	+
	1980Q1	Carter–Reagan build-up	+
	2001Q4	Post-9/11 build-up	+

Notes: Panel (a) reports the sign restrictions; each entry is imposed on impact and over the following three quarters (4q), and a blank cell is left unrestricted. The system contains five variables; the Treasury bill rate and the GDP deflator are unrestricted for both fiscal shocks. Panel (b) reports the narrative restrictions. At each episode the realized fiscal shock carries the indicated sign. Its absolute contribution to the unexpected change in its own policy variable exceeds the sum of the absolute contributions of all other shocks. We identify both fiscal shocks simultaneously in the five-variable system, so each carries its own sign and narrative restrictions.

TABLE 9. Dates for variance regimes (1947Q1–2007Q4), 3-regime specification

Regime	Start	End	Description
1	1947Q1	1971Q1	Bretton Woods era, including the Korean War mobilization.
2	1971Q2	1996Q4	End of Bretton Woods through the Great Inflation and Volcker disinflation.
3	1997Q1	2007Q4	Late-1990s deficit reduction and pre-crisis period.

Notes: Break dates follow CMT (App. D). End dates are inclusive. Both break dates are distinct from the narrative episode quarters (1951Q1, 1964Q2, 1965Q3, 1980Q1, 1993Q4, and 2001Q4).

TABLE 10. Effect of posterior-draw repair across applications

Application	N_θ	Hours/1,000 draws	Stale-Q events	Repaired (%)
Monetary	100	0.097	9,165	98.0
	0	0.049	8,739	0.0
Oil	100	0.034	4,181	69.3
	0	0.006	3,744	0.0
Fiscal	100	0.035	14,941	91.7
	0	0.008	14,461	0.0

Notes: Each application uses 100,000 saved draws of (θ, Q) after burn-in, including the rotations saved within each parameter iteration. Timing is measured in hours per 1,000 effective impulse response draws, computed as in Table S4. A stale-Q event occurs when a model-parameter proposal makes the current rotation Q inadmissible. When $N_\theta = 100$, HARS attempts posterior-draw repair: holding Q fixed, it makes up to 100 further completed draws of (A_0, Λ, A_+), each starting from the last failed candidate, and retains the first admissible one (Theorem 1(ii)). In the oil application only (A_0, Λ) enters the admissibility check, because the restrictions bind only on impact. The $N_\theta = 0$ rows disable the repair. *Repaired* is the share of stale-Q events resolved by the repair. An unrepaired event leaves the parameter chain at its current admissible draw; it affects mixing, not the target.

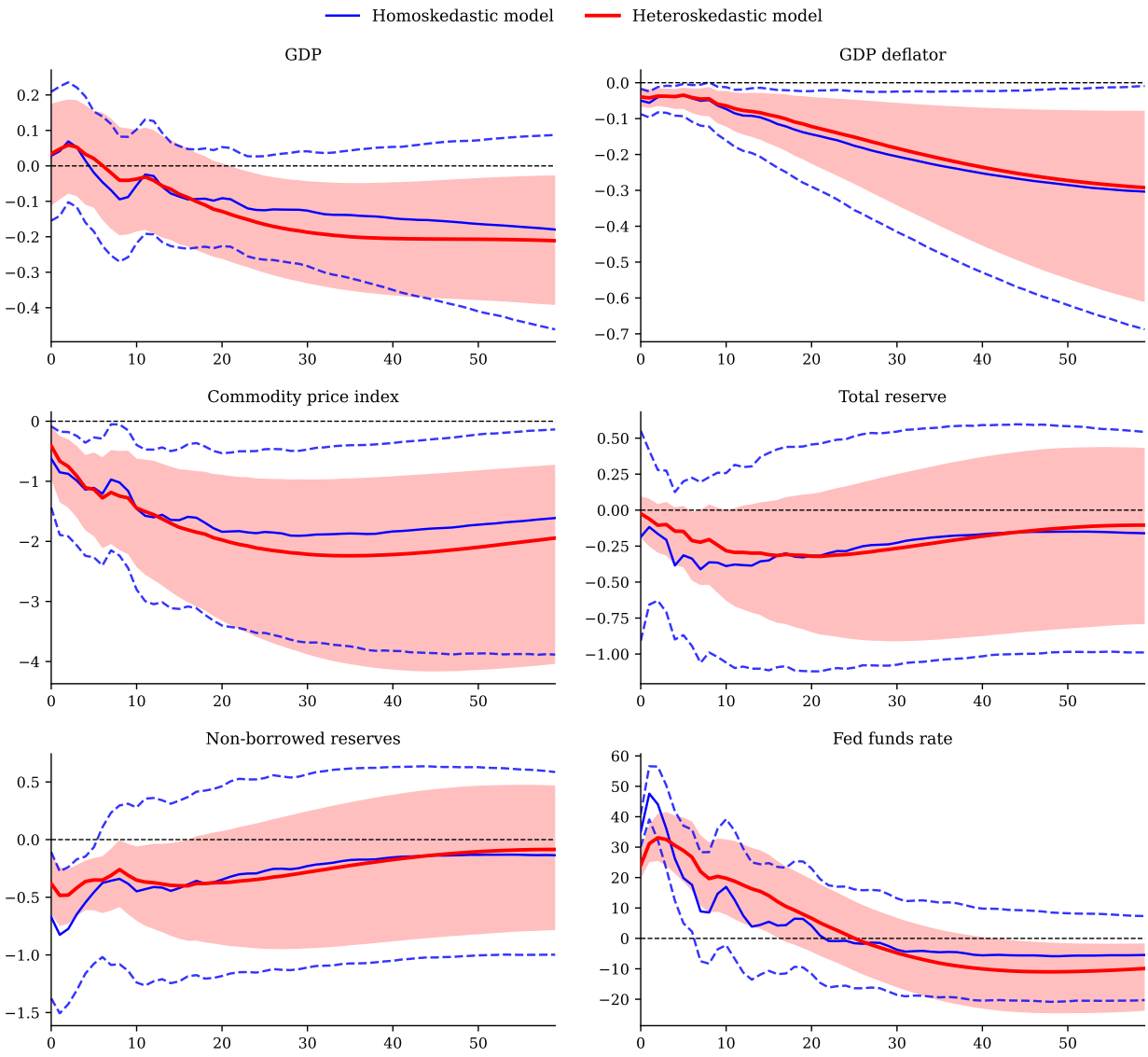


FIGURE 1. Impulse responses to a contractionary monetary policy shock: homoskedastic specification (ADRR) versus heteroskedastic specification. Solid lines are posterior medians; bands are 90% credible intervals. Vertical axes are in percent, except for the federal funds rate, which is in basis points. Horizontal axes are months after the shock.

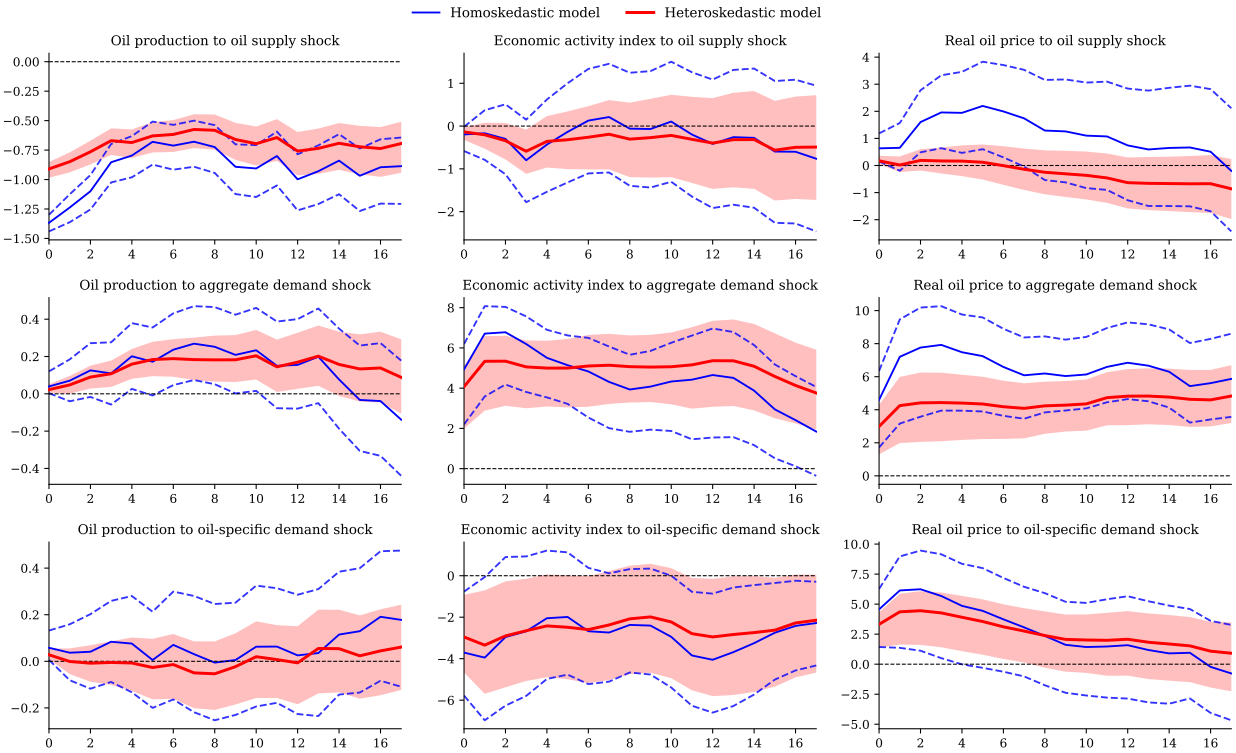


FIGURE 2. Impulse responses to the three oil market shocks: homoskedastic specification (KM) versus heteroskedastic specification. KM's homoskedastic specification is shown in blue and is estimated using the ADRR codebase; the heteroskedastic specification is shown in red. Solid lines are posterior medians; bands are 90% credible intervals. Vertical axes are in percent, and horizontal axes are months after the shock.

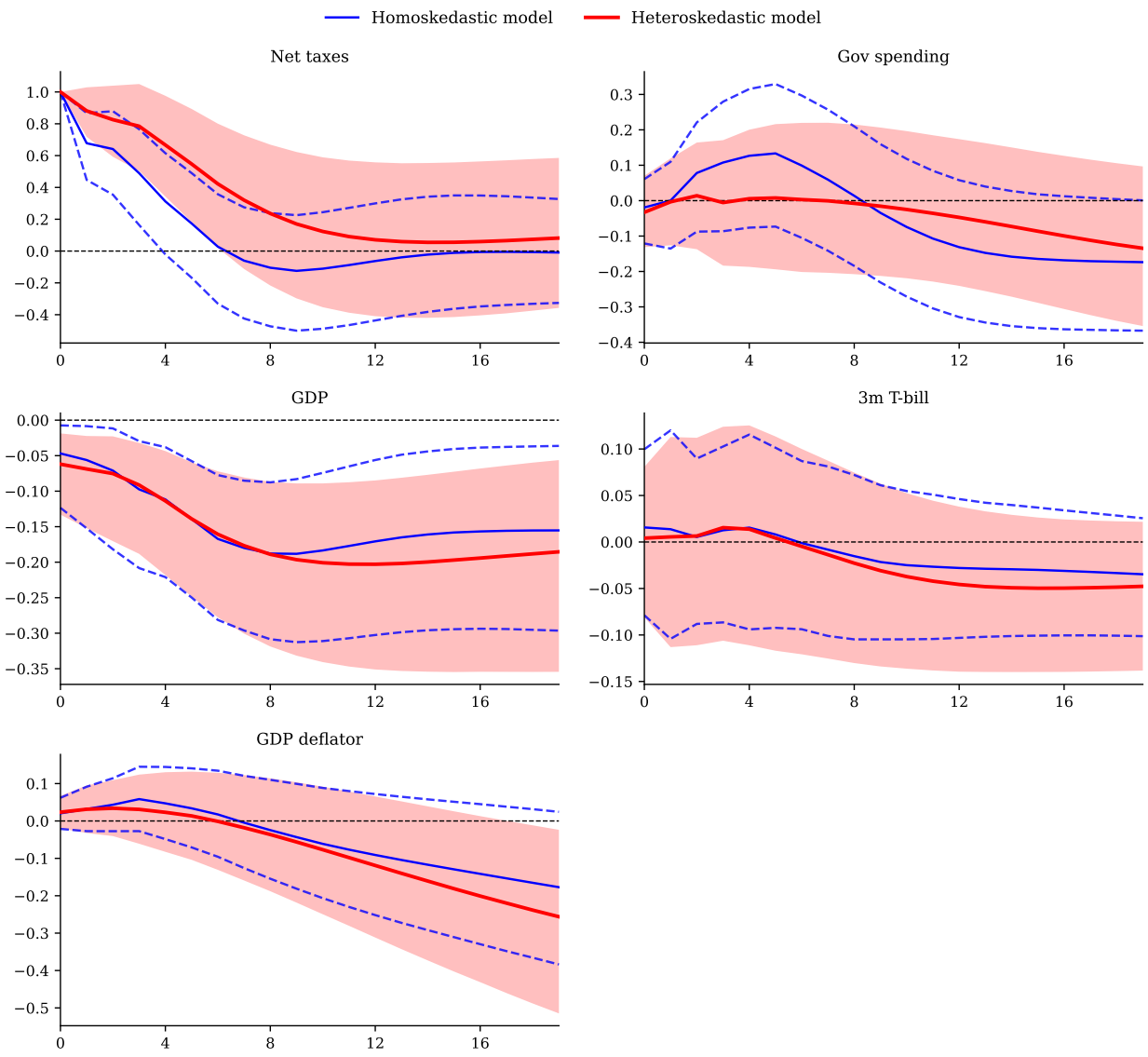


FIGURE 3. Responses to a net-tax shock normalized in each posterior draw to raise real per-capita net taxes by 1% on impact. Blue denotes the homoskedastic specification and red the heteroskedastic specification. Solid lines are posterior medians; bands are pointwise equal-tailed 90% intervals. Quantity and price responses are in percent; the Treasury-bill-rate response is in percentage points. Horizontal axes are quarters after the shock, with impact at quarter 0.

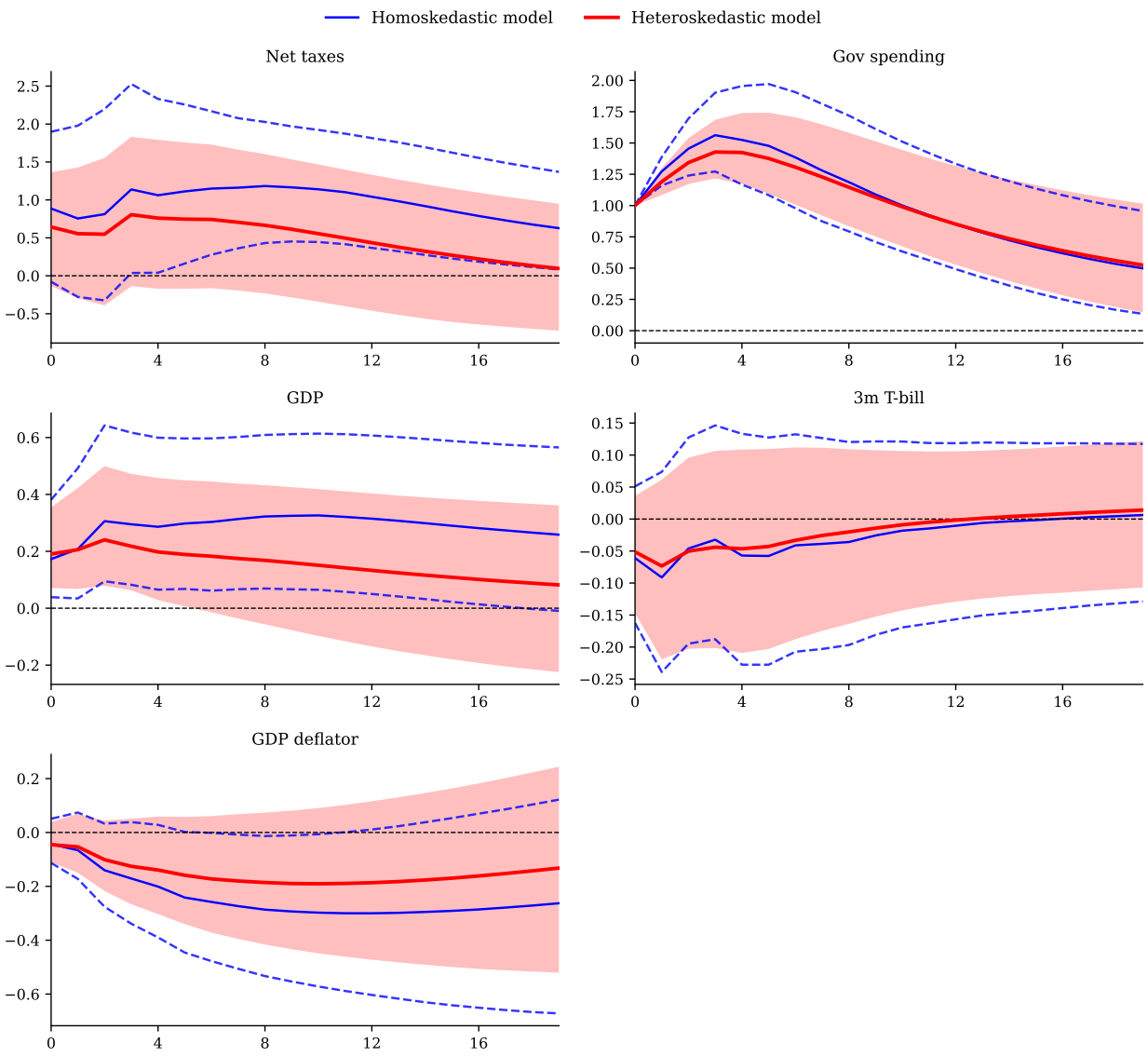


FIGURE 4. Responses to a government spending shock normalized in each posterior draw to raise real per-capita government spending by 1% on impact. Blue denotes the homoskedastic specification and red the heteroskedastic specification. Solid lines are posterior medians; bands are pointwise equal-tailed 90% intervals. Quantity and price responses are in percent; the Treasury-bill-rate response is in percentage points. Horizontal axes are quarters after the shock, with impact at quarter 0.

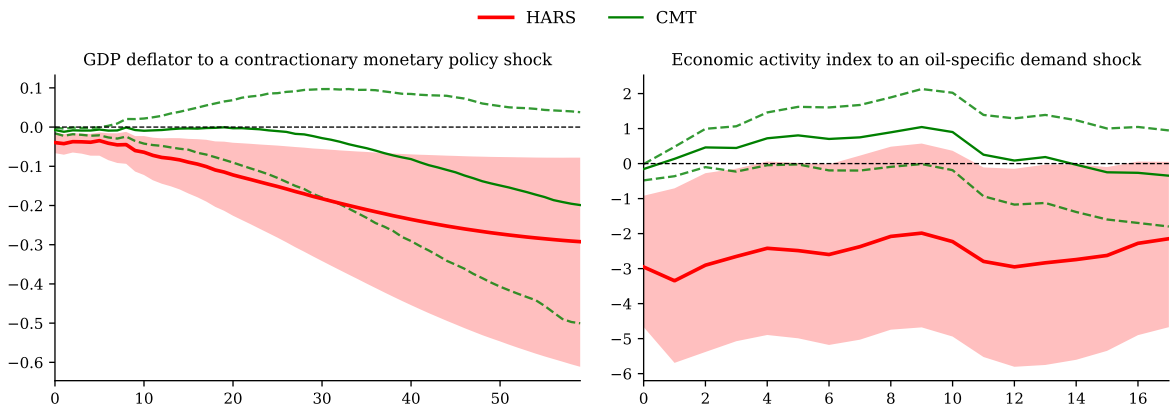


FIGURE 5. Impulse responses to a contractionary monetary policy shock (left panel) and a positive oil-specific demand shock (right panel): HARS (red) versus CMT's Algorithm 2 (green). Vertical axes are in percent, and horizontal axes are months after the shock.

In this appendix, sections, tables, figures, and equations are labeled with the prefix "S" to denote supplemental material.

Appendix S1. Proofs

This section provides the proofs of the main-text lemma, propositions, and theorem, together with the supporting propositions on the sign-restricted target, ESS, and rejection counts.

S1.1. **Proof of Lemma 1.** The vector η has mean zero and covariance matrix Λ_k. Since $\widetilde{\eta}' = \eta' Q$, the corresponding column vector is

$$\widetilde{\eta} = Q'\eta.$$

Hence $\widetilde{\eta}$ has mean zero and covariance matrix

$$\mathrm{Var}(\widetilde{\eta}) = Q' \,\mathrm{Var}(\eta) Q = Q'\Lambda_k Q.$$

The components of $\widetilde{\eta}$ are mutually uncorrelated if and only if this covariance matrix is diagonal. Under joint Gaussianity, this is also equivalent to mutual independence. Conditioning only on the regime does not give Gaussianity under the maintained scale-mixture specification, so the general assertion concerns covariance. This proves the first claim.

It remains to characterize when $Q'\Lambda_k Q$ is diagonal. Suppose first that the diagonal entries of Λ_k are distinct. Let

$$D = Q'\Lambda_k Q.$$

If D is diagonal, then multiplying by Q on the left gives

$$\Lambda_k Q = QD.$$

Let q_j denote the jth column of Q, and let d_j denote the jth diagonal entry of D. The jth column of the last display gives

$$\Lambda_k q_j = d_j q_j.$$

Thus every column of Q must be an eigenvector of Λ_k. Since Λ_k is diagonal with distinct diagonal entries, each eigenspace is one-dimensional and is spanned by one coordinate vector. Therefore each column of Q must be a coordinate vector up to sign:

$$q_j = \pm e_{\pi(j)}$$

for some map π from $\{1, \ldots, n\}$ into itself. Orthogonality of Q rules out repeated coordinate vectors, so π is a permutation and Q is a signed permutation matrix.

Conversely, if Q is a signed permutation matrix, then $Q'\Lambda_k Q$ only permutes the diagonal entries of Λ_k. The result is still diagonal, so diagonal shock covariance is preserved. This proves the second claim in the distinct-eigenvalue case.

If Λ_k has repeated diagonal entries, the same eigenvector argument applies after grouping equal entries. On any block with a common diagonal value, Λ_k is proportional to the identity,

so an orthogonal rotation within that block leaves the block diagonal. An orthogonal rotation that mixes two blocks with different diagonal values cannot preserve diagonality, because it would mix eigenspaces associated with different eigenvalues. Hence, up to a permutation of columns, the covariance-preserving rotations are orthogonal rotations within equal-variance eigenspaces. In particular, when the regime-specific shock variances are unequal across shocks, a generic orthogonal rotation does not preserve diagonal shock covariance.

□

S1.2. **Proof of Proposition 1.** Write $W_k = A_0\Xi_k$ and $e_t = Q'\varepsilon_t$. The rotated coefficient block is W_kQ, and the reduced-form coefficients satisfy

$$(A_+\Xi_kQ)(A_0\Xi_kQ)^{-1} = A_+A_0^{-1} = B.$$

If $f_{\varepsilon,t}$ is the joint density of the standardized model shock, possibly conditional on the latent scales, the induced density of e_t is

$$f_{e,t}(e \mid Q) = f_{\varepsilon,t}(Qe),$$

because $|\det Q| = 1$. At the observed residual,

$$e_t = Q'\Xi_k(A_0'y_t - A_+'x_t),$$

so

$$f_{e,t}(e_t \mid Q) = f_{\varepsilon,t}\big(\Xi_k(A_0'y_t - A_+'x_t)\big).$$

The Jacobian is unchanged as well: $|\det(W_kQ)| = |\det W_k|$. Multiplying the density and Jacobian terms over dates proves $p(Y \mid \theta, Q) = p(Y \mid \theta)$.

Conditional on the latent scales, the rotated shock covariance is $Q'\operatorname{diag}(\nu_t)Q$; it is not generally diagonal. The argument transforms that full Gaussian distribution rather than assigning a fresh independent shock law after rotation. Integration over the latent scales therefore preserves the likelihood equality. At the covariance level the impact matrix $S_k = (A_0^{-1})'\Lambda_k^{1/2}Q$ satisfies

$$S_kS_k' = (A_0^{-1})'\Lambda_kA_0^{-1} = \Sigma_k.$$

Thus both the covariance and the full data distribution are unchanged under the induced shock law. This proves the proposition. □

S1.3. **Proof of Proposition 2.** Under the coefficient-only rotation, the reduced-form coefficients remain B, but retaining the original shock law gives covariance

$$\widetilde{\Sigma}_k(Q) = (A_0^{-1})'Q\Lambda_kQ'A_0^{-1}.$$

Since A_0 is nonsingular, equality with Σ_k requires $Q\Lambda_kQ' = \Lambda_k$. Equal likelihoods for all data imply equal data distributions and hence equal covariances, so this is a necessary condition. It fails for a generic rotation when Λ_k is not scalar. For Gaussian shocks without equation-specific latent scales, the mean and covariance determine the distribution, so the condition

is also sufficient. For scale-mixture shocks, covariance equality alone need not preserve the full shock density. This proves the stated necessary condition. □

S1.4. **The sign-restricted target.**

Proposition S1 (Conditional structure of the target)**.** *Under Definition 1:*

(i) conditional on θ, Q has Haar measure restricted to $R(\theta)$;
(ii) conditional on Q, θ has density proportional to $p(\theta \mid Y)\,\mathbf{1}\{SR(\theta, Q) > 0\}$;
(iii) the marginal density of θ is $p(\theta \mid Y)\,v(\theta)/c(Y)$.

Consequently, a transition that updates θ at fixed Q with a kernel leaving (ii) invariant, or updates Q at fixed θ with a kernel leaving (i) invariant, leaves the target of Definition 1 invariant. Neither update requires $v(\theta)$.

Proof. By Proposition 1, $p(Y \mid \theta, Q) = p(Y \mid \theta)$ for every Q. Before the restrictions are imposed, the relevant product measure for (θ, Q) is therefore

$$p(\theta \mid Y)\, d\theta\, \mu(dQ),$$

where μ is Haar measure on $\mathcal{O}(n)$. The data update θ through the heteroskedastic likelihood, while Q enters only through the restriction check. Definition 1 multiplies this product measure by $\mathbf{1}\{Q \in R(\theta)\}$ and divides by $c(Y)$.

For (i), fix θ with $v(\theta) > 0$. As a measure in Q, the target is proportional to $\mathbf{1}\{Q \in R(\theta)\}\mu(dQ)$. Its normalized version is $\mathbf{1}\{Q \in R(\theta)\}\mu(dQ)/v(\theta)$, Haar measure restricted to $R(\theta)$. For (ii), fix Q. As a function of θ, the target density is proportional to $p(\theta \mid Y)\mathbf{1}\{SR(\theta, Q) > 0\}$. For (iii), integrate over Q:

$$\int \frac{p(\theta \mid Y)\,\mathbf{1}\{Q \in R(\theta)\}}{c(Y)}\,\mu(dQ) = \frac{p(\theta \mid Y)\,v(\theta)}{c(Y)}.$$

For the final claim, write the target as the marginal of θ times the conditional of Q given θ. A transition that holds θ fixed and moves Q with a kernel leaving (i) invariant preserves the marginal of θ and, for each θ, the conditional of Q; hence it preserves the joint target. The same argument, with the roles of θ and Q interchanged and (ii) in place of (i), applies to a parameter update at fixed Q. The argument carries over to the Gaussian coordinates of Proposition S2: the joint density proportional to $p(\theta \mid Y)\,\varphi(X)\,\mathbf{1}\{SR(\theta, \gamma(X)) > 0\}$, with φ the standard Gaussian density, has image Definition 1 under $(\theta, X) \mapsto (\theta, \gamma(X))$, because $\gamma(X)$ is Haar distributed when X is standard Gaussian. Augmenting θ with latent scales that do not enter the restrictions leaves the argument unchanged. Finally, a rejection step over rotations at a fixed θ implements (i). It does not by itself deliver the marginal (iii) of θ, which requires a parameter update that conditions on the rotation. This proves the proposition.

□

S1.5. **Admissible rotations in Gaussian coordinates.**

Proposition S2. *Fix θ with $\mu(R(\theta)) > 0$. Let $X = (x_1, \dots, x_n)$ have independent standard-normal columns and let $Q = \gamma(X)$. Define the sign-restricted distribution of X by*

$$\pi_X(dX \mid \theta, SR) \propto \mathbf{1}\{SR(\theta, \gamma(X)) > 0\} \exp\Big(-\tfrac{1}{2}\textstyle\sum_{j=1}^{n} \|x_j\|^2\Big)\, dX.$$

Then the image of $\pi_X(\cdot \mid \theta, SR)$ under γ is Haar measure restricted to $R(\theta)$—the conditional target density of Definition 1. The full conditional density of a single column is

$$p(x_j \mid x_{-j}, \theta, SR) \propto \mathbf{1}\{SR(\theta, \gamma(X)) > 0\} \exp\big(-\tfrac{1}{2}\|x_j\|^2\big)\,,$$

a Gaussian density times an indicator. A column-wise ESS update of x_j leaves this full conditional density invariant, and a systematic scan over $j = 1, \dots, n$ preserves the invariance of $\pi_X(\cdot \mid \theta, SR)$.

Proof. Let X be an $n \times n$ matrix with independent standard-normal entries, and let $Q = \gamma(X)$ be the orthogonal factor obtained from the QR construction with the positive-diagonal convention. The set of rank-deficient X has Gaussian probability zero, so the map is well defined almost surely. The standard Gaussian distribution is invariant under orthogonal transformations, and the QR normalization removes the sign ambiguity. Therefore $\gamma(X)$ is Haar distributed on $\mathcal{O}(n)$. In particular, for every measurable $C \subseteq \mathcal{O}(n)$,

$$\Pr\{\gamma(X) \in C\} = \mu(C).$$

The sign-restricted density of X is the standard Gaussian density multiplied by the indicator $\mathbf{1}\{SR(\theta, \gamma(X)) > 0\}$. The image of this density under γ is therefore obtained by conditioning the induced Haar draw on the event $Q \in R(\theta)$. Hence, for every measurable $C \subseteq \mathcal{O}(n)$,

$$\Pr\{\gamma(X) \in C \mid SR(\theta, \gamma(X)) > 0\} = \frac{\mu(C \cap R(\theta))}{\mu(R(\theta))}.$$

This is exactly Haar measure restricted to $R(\theta)$, which is the conditional target for Q in Definition 1. This proves the image claim.

Next fix the columns x_{-j}. As a function of x_j, the joint density of X factors into a standard normal density in x_j, a term depending only on x_{-j}, and the admissibility indicator. The full conditional density is therefore

$$p(x_j \mid x_{-j}, \theta, SR) \propto \mathbf{1}\{SR(\theta, \gamma(X)) > 0\} \exp\big(-\tfrac{1}{2}\|x_j\|^2\big)\,,$$

which is a Gaussian density times an indicator.

Elliptical slice sampling is exact for a Gaussian density multiplied by a nonnegative likelihood factor (Murray et al., 2010); here the likelihood factor is the admissibility indicator. Draw an auxiliary vector $\tau_j \sim \mathcal{N}(0, I_n)$ and consider the ellipse

$$x_j(\vartheta) = x_j \cos\vartheta + \tau_j \sin\vartheta, \qquad \vartheta \in [0, 2\pi].$$

The joint standard-normal density of (x_j, τ_j) is invariant under this two-dimensional rotation. Therefore moving along the ellipse preserves the Gaussian part of the full conditional exactly. The slice step retains only angles for which

$$SR(\theta, \gamma(X(\vartheta))) > 0.$$

Since the current point is admissible, the admissible angular set is nonempty. The bracket-shrinking rule samples from this set in a way that satisfies detailed balance for the full conditional. Thus the column update leaves $p(x_j \mid x_{-j}, \theta, SR)$ invariant.

A systematic scan over the columns composes invariant full-conditional updates, so it leaves $\pi_X(\cdot \mid \theta, SR)$ invariant. Mapping the resulting X through γ gives a draw of Q whose invariant distribution is Haar measure restricted to $R(\theta)$. This proves the proposition.

□

S1.6. **Proof of Theorem 1.** Throughout, all integrands are nonnegative, so the order of integration may be exchanged freely and every posterior density below is well defined. We argue in five steps.

Step 1: the target and its z_0-marginal. The unrestricted completed draw has density

$$p_0(z_0 \mid s_0)\, p_+(z_+ \mid z_0, s_0),$$

the first block times the companion conditional. Imposing sign admissibility multiplies this density by the indicator $\mathbf{1}\{SR(z_0, z_+, Q) > 0\}$ and renormalizes. This gives the stated target $\pi^{SR}(z_0, z_+ \mid s_0, Q)$. To obtain its z_0-marginal, integrate the target over z_+:

$$\pi_0^{SR}(z_0 \mid s_0, Q) \propto p_0(z_0 \mid s_0) \int p_+(z_+ \mid z_0, s_0)\, \mathbf{1}\{SR(z_0, z_+, Q) > 0\}\, dz_+.$$

The inner integral is exactly $\omega_0(z_0; s_0, Q)$, the probability that a companion draw from $p_+(\cdot \mid z_0, s_0)$ lands in the admissible region for the given z_0 and Q. Hence

$$\pi_0^{SR}(z_0 \mid s_0, Q) \propto p_0(z_0 \mid s_0)\, \omega_0(z_0; s_0, Q),$$

which is the marginal claim. The weight ω_0 is the entire difference between the unrestricted marginal p_0 and the sign-restricted marginal. A first block z_0 whose admissible completions are rare carries a small ω_0 and must therefore receive proportionally less mass under the target.

Step 2: what the conditional procedure samples. For part (i), the completion probability is positive for p_0-almost every z_0, and the companion-only search therefore terminates almost surely. Fix such a z_0. If this positivity condition fails, the search never terminates at a value with $\omega_0(z_0; s_0, Q) = 0$. Draw companion candidates $u \sim p_+(\cdot \mid z_0, s_0)$ and accept the first one that satisfies $SR(z_0, u, Q) > 0$. Each candidate is drawn from the same density and

is retained exactly on the admissible event. By the standard accept–reject argument, the accepted companion has density

$$p_+^{SR}(z_+ \mid z_0, s_0, Q) = \frac{p_+(z_+ \mid z_0, s_0)\,\mathbf{1}\{SR(z_0, z_+, Q) > 0\}}{\omega_0(z_0; s_0, Q)}.$$

This is a correct conditional density: it is nonnegative and, by the definition of ω_0, integrates to one over z_+. Its normalizing constant is $\omega_0(z_0; s_0, Q)$.

Now compose the full conditional procedure. Draw $z_0 \sim p_0(\cdot \mid s_0)$, then draw z_+ by the accept–reject step just described. The retained pair has joint density

$$p_0(z_0 \mid s_0)\, p_+^{SR}(z_+ \mid z_0, s_0, Q) = p_0(z_0 \mid s_0)\, \frac{p_+(z_+ \mid z_0, s_0)\,\mathbf{1}\{SR(z_0, z_+, Q) > 0\}}{\omega_0(z_0; s_0, Q)}.$$

Integrate out z_+. Because $p_+^{SR}(\cdot \mid z_0, s_0, Q)$ integrates to one for every fixed z_0,

$$\int p_0(z_0 \mid s_0)\, p_+^{SR}(z_+ \mid z_0, s_0, Q)\, dz_+ = p_0(z_0 \mid s_0).$$

The z_0-marginal is therefore $p_0(z_0 \mid s_0)$, not $p_0(z_0 \mid s_0)\,\omega_0(z_0; s_0, Q)$. The mechanism of the error is explicit: $\omega_0(z_0; s_0, Q)$ is both the normalizing constant that makes the companion conditional integrate to one and the factor that the target places on z_0. Dividing by it to form the conditional density removes it from the marginal. The conditional procedure therefore samples the companion block correctly, given z_0, but strips the first block of exactly the admissibility weight the target requires. This proves (i).

Step 3: the completed-draw procedure with independent candidates. Consider first completed candidates whose first block is drawn independently from $p_0(\cdot \mid s_0)$, with $z_+ \sim p_+(\cdot \mid z_0, s_0)$, retaining the first candidate that satisfies $SR(z_0, z_+, Q) > 0$. The candidate density is $p_0(z_0 \mid s_0)\, p_+(z_+ \mid z_0, s_0)$, and retention occurs with probability $\mathbf{1}\{SR(z_0, z_+, Q) > 0\}$. The overall retention probability is

$$\begin{aligned}\alpha &= \iint p_0(z_0 \mid s_0)\, p_+(z_+ \mid z_0, s_0)\,\mathbf{1}\{SR(z_0, z_+, Q) > 0\}\, dz_+\, dz_0 \\ &= \int p_0(z_0 \mid s_0)\,\omega_0(z_0; s_0, Q)\, dz_0.\end{aligned}$$

This probability is positive whenever the admissible region has positive probability under the candidate distribution; this is the maintained condition under which the repair is invoked. By the accept–reject argument, the retained pair has joint density

$$\frac{p_0(z_0 \mid s_0)\, p_+(z_+ \mid z_0, s_0)\,\mathbf{1}\{SR(z_0, z_+, Q) > 0\}}{\alpha} = \pi^{SR}(z_0, z_+ \mid s_0, Q),$$

the target itself. Its z_0-marginal follows by integrating out z_+:

$$\frac{1}{\alpha}\, p_0(z_0 \mid s_0) \int p_+(z_+ \mid z_0, s_0)\,\mathbf{1}\{SR(z_0, z_+, Q) > 0\}\, dz_+ = \frac{1}{\alpha}\, p_0(z_0 \mid s_0)\,\omega_0(z_0; s_0, Q).$$

This marginal is proportional to $p_0(z_0 \mid s_0)\,\omega_0(z_0; s_0, Q)$, the correct sign-restricted marginal from Step 1. With the independence proposal $q_0(\cdot \mid z_0) = p_0(\cdot \mid s_0)$, the MH probability

in (ii) is $\alpha_{\mathrm{MH}} = 1$, and each completed transition draws independently from $p_0 p_+$. The calculation above gives the distribution of an admissible candidate. Under the fixed budget in (ii), a search that finds no admissible candidate retains the original pair; Step 5 proves invariance with this outcome included. Here the admissibility weight is retained rather than divided out, because the candidate z_0 must itself survive the sign check together with its companion.

Step 4: when the two procedures agree. Under the positivity condition in part (i), compare the two retained z_0-marginals. The conditional procedure yields $p_0(z_0 \mid s_0)$. The completed-draw procedure yields

$$p_0(z_0 \mid s_0)\, \omega_0(z_0; s_0, Q)/\alpha.$$

These two densities are equal for p_0-almost every z_0 if and only if

$$\omega_0(z_0; s_0, Q) = \alpha \quad \text{for } p_0\text{-almost every } z_0,$$

that is, if and only if $\omega_0(z_0; s_0, Q)$ is constant in z_0 on the support of p_0. Generically it is not: the probability that a companion draw is admissible varies with the first block, so ω_0 varies with z_0. Outside that knife-edge case, the conditional procedure misrepresents the marginal density of z_0, and only the completed-draw repair reproduces the target.

Step 5: the MH-first repair from the last failed candidate. Write

$$f(d\theta) = p_0(dz_0 \mid s_0)\, p_+(dz_+ \mid z_0, s_0).$$

The ordinary first-block MH transition satisfies

$$p_0(da \mid s_0) K_0(a, da') = p_0(da' \mid s_0) K_0(a', da).$$

For completed states $x = (a, b)$ and $y = (a', b')$, define

$$C(x, dy) = K_0(a, da')\, p_+(db' \mid a', s_0).$$

This transition includes a new companion draw after a first-block MH rejection. It is reversible for f, because

$$\begin{aligned} f(dx)C(x, dy) &= p_0(da \mid s_0) K_0(a, da') p_+(db \mid a, s_0) p_+(db' \mid a', s_0) \\ &= p_0(da' \mid s_0) K_0(a', da) p_+(db' \mid a', s_0) p_+(db \mid a, s_0) \\ &= f(dy)C(y, dx). \end{aligned}$$

Let

$$S = \{\theta : SR(\theta, Q) > 0\}, \qquad Z = f(S) > 0, \qquad \pi^{SR}(d\theta \mid s_0, Q) = \frac{\mathbf{1}_S(\theta) f(d\theta)}{Z}.$$

Starting from the original admissible state $x_0 \in S$, generate successive states x_κ from $C(x_{\kappa-1}, \cdot)$. Steps 1–2 of Algorithm 1 generate at most $N = N_\theta + 1$ such transitions. Each transition after a failure starts from that failed state. If the first return to S occurs within this budget, retain it; otherwise retain x_0.

For a first-return path of length $\nu \le N$, the endpoints x_0, x_ν lie in S and every intermediate state $x_1, \dots, x_{\nu-1}$ lies outside S. Reversibility of C gives equality of the forward and reverse path measures:

$$\begin{aligned} &f(dx_0)C(x_0, dx_1)\cdots C(x_{\nu-1}, dx_\nu) \\ &\quad = f(dx_\nu)C(x_\nu, dx_{\nu-1})\cdots C(x_1, dx_0). \end{aligned}$$

Reversing the path preserves its length and the failure of every intermediate state. Divide by Z, integrate over the intermediate states outside S, and sum over $\nu = 1, \dots, N$. This gives detailed balance for all moves between distinct admissible endpoints. Paths that fail to return within the budget place their probability on retaining x_0, which also satisfies detailed balance. Thus the finite-budget transition is reversible for $\pi^{SR}(\cdot \mid s_0, Q)$. The ordinary MH decisions inside K_0 provide the required correction; no endpoint acceptance test or product over failed candidates is needed. This proves (ii).

□

S1.7. **Rejection counts.**

Proposition S3 (Rejection counts carry no likelihood content)**.** *Fix a draw of θ with $v(\theta) > 0$. Draw $Q^{(1)}, Q^{(2)}, \dots$ independently from Haar measure and retain the first that satisfies $SR(\theta, Q) > 0$. Then:*

- *(i) the retained rotation has the restricted-Haar distribution of Proposition S1(i);*
- *(ii) the likelihood contribution of θ is unchanged;*
- *(iii) the realized rejection rate of any rotation search or repair step depends on the conditioning state and on the proposal mechanism. It is a diagnostic of computational cost. The admissible volume $v(\theta)$ in Proposition S1(iii) is accounted for by the admissibility checks in the parameter update, so rejection counts must not be used as weights.*

Proof. For (i), let

$$J = \min\{m : SR(\theta, Q^{(m)}) > 0\}$$

be the first admissible proposal. The proposals are independent Haar draws. For any measurable $C \subseteq \mathcal{O}(n)$,

$$\begin{aligned} \Pr\{Q^{(J)} \in C\} &= \sum_{m=1}^{\infty} \Pr\{Q^{(1)}, \dots, Q^{(m-1)} \notin R(\theta),\ Q^{(m)} \in C \cap R(\theta)\} \\ &= \sum_{m=1}^{\infty} \{1 - \mu(R(\theta))\}^{m-1} \mu(C \cap R(\theta)) \\ &= \frac{\mu(C \cap R(\theta))}{\mu(R(\theta))}, \end{aligned}$$

where the last equality uses $\mu(R(\theta)) > 0$. Hence the retained rotation has density $\mathbf{1}\{Q \in R(\theta)\}/\mu(R(\theta))$ against Haar measure, which is the conditional distribution in Proposition S1(i).

For (ii), Proposition 1 gives

$$p(Y \mid \theta, Q) = p(Y \mid \theta)$$

for every Q. Therefore replacing an inadmissible rotation by an admissible one changes the impulse responses computed from (θ, Q), but not the likelihood contribution of θ.

For (iii), the relevant admissibility probabilities are conditional on the state and on the repair mechanism. For a companion-block update with z_0 fixed, the per-candidate admissibility probability is

$$\alpha_+(z_0; s_0, Q) = \int p_+(z_+ \mid z_0, s_0)\, \mathbf{1}\{SR(z_0, z_+, Q) > 0\}\, dz_+.$$

For a completed-draw candidate generated by the MH-first transition from z_0, it is

$$\alpha_0(z_0; s_0, Q) = \int \omega_0(a; s_0, Q)\, K_0(z_0, da).$$

This expression includes first-block MH rejection followed by a new companion draw.

These probabilities are not likelihood terms. They are probabilities of finding an admissible proposal under the proposal mechanism being used, and they change with the conditioning state, the current rotation, and the block being updated. A realized aggregate rejection rate averages over different states and mechanisms, so it is not a state-specific normalizing constant. The weight $v(\theta)$ of Proposition S1(iii) does enter the target, but the parameter update of Theorem 1(ii) realizes it through admissibility checks without estimating it. Inserting rejection counts as additional weights would therefore distort the target. Rejection counts are diagnostics of computational efficiency, not density adjustments. This proves the proposition.

Appendix S2. Complete derivation of Example 1

Let P_H and P_O denote probabilities under the respective restricted-Haar laws for both models, conditional on the specified parameters, where the subscripts H and O refer to the heteroskedastic and homoskedastic models.

Step 1: Specify the model and compute its covariances. There are two variables, no lags, and zero deterministic terms:

$$y_t' A_0 = \eta_t', \qquad A_0 = \begin{pmatrix} -2/3 & 5/3 \\ 5/3 & -5/3 \end{pmatrix}, \qquad A_0^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 2/5 \end{pmatrix}.$$

Columns of A_0 correspond to equations:

$$\frac{-2y_{1,t} + 5y_{2,t}}{3} = \eta_{1,t}, \qquad \frac{5y_{1,t} - 5y_{2,t}}{3} = \eta_{2,t}.$$

Equivalently, $y_{1,t} = \eta_{1,t} + \eta_{2,t}$ and $y_{2,t} = \eta_{1,t} + (2/5)\eta_{2,t}$. Since the conditional mean is zero,

$$u_t = y_t, \qquad \eta_t = A_0' u_t = \begin{pmatrix} (-2u_{1,t} + 5u_{2,t})/3 \\ (5u_{1,t} - 5u_{2,t})/3 \end{pmatrix}.$$

The regimes are known and equally frequent. Independently across dates, $\eta_t \mid s_t = k \sim \mathcal{N}(0, \Lambda_k)$, where

$$\Lambda_1 = \mathrm{diag}(8/5, 2/5), \qquad \Lambda_2 = \mathrm{diag}(2/5, 8/5).$$

The average $(\Lambda_1 + \Lambda_2)/2 = I_2$. Direct multiplication gives

$$\Sigma_1 = (A_0^{-1})' \Lambda_1 A_0^{-1} = \begin{pmatrix} 2 & 44/25 \\ 44/25 & 208/125 \end{pmatrix},$$

$$\Sigma_2 = (A_0^{-1})' \Lambda_2 A_0^{-1} = \begin{pmatrix} 2 & 26/25 \\ 26/25 & 82/125 \end{pmatrix}.$$

Step 2: Recover the shocks and define the rotation. The heteroskedastic model recovers

$$\varepsilon_t = \Lambda_{s_t}^{-1/2} A_0' u_t.$$

The homoskedastic model recovers $\varepsilon_t = A_0' u_t$. In each model the notation refers to its own recovered standardized shocks. Write

$$Q = \begin{pmatrix} \cos\psi & -\iota \sin\psi \\ \sin\psi & \iota \cos\psi \end{pmatrix}, \qquad -\pi < \psi \le \pi, \qquad \iota \in \{-1, 1\}.$$

Under Haar measure, ψ is uniform and ι is an independent equally likely sign. The rotated first shock is

$$e_{1,t} = \varepsilon_{1,t} \cos\psi + \varepsilon_{2,t} \sin\psi.$$

The response experiment sets $e_t = (1, 0)'$, holding the other shock at zero.

Step 3: Check the same response sign in every regime. The reported first-shock impact column is

$$(A_0^{-1})' q_1 = \begin{pmatrix} \cos\psi + \sin\psi \\ \cos\psi + \frac{2}{5}\sin\psi \end{pmatrix}.$$

Require only a positive response of y_1:

$$\cos\psi + \sin\psi > 0.$$

Taking the representative angle interval $(-\pi/4, 3\pi/4)$, this is the entire homoskedastic admissible set. Repeating the same check in different dates does not change it.

In regime 1, the heteroskedastic response of y_1 is

$$\sqrt{8/5}\cos\psi + \sqrt{2/5}\sin\psi = \sqrt{2/5}(2\cos\psi + \sin\psi).$$

In regime 2 it is

$$\sqrt{2/5}\cos\psi + \sqrt{8/5}\sin\psi = \sqrt{2/5}(\cos\psi + 2\sin\psi).$$

With $\gamma = \arctan(1/2)$, the first of these inequalities holds on $(-\pi/2 + \gamma, \pi/2 + \gamma)$, and the second holds on $(-\gamma, \pi - \gamma)$, modulo 2π. Intersecting them with the angle interval satisfying the restriction on the reported response gives

$$\psi \sim \text{Uniform}(-\gamma, \pi/2 + \gamma) \quad \text{under the heteroskedastic model,}$$
$$\psi \sim \text{Uniform}(-\pi/4, 3\pi/4) \quad \text{under the homoskedastic model.}$$

The additional regime checks exclude a set of positive Haar probability. No narrative restriction is imposed.

The response of interest, with no sign restriction on it, is

$$\Theta_{21,0} = \cos\psi + \tfrac{2}{5}\sin\psi.$$

This formula is identical under the two models; only the admitted angles differ.

Step 4: Compute the probabilities and every reported quantile. Set

$$R = \frac{\sqrt{29}}{5}, \qquad \delta = \arctan(2/5), \qquad \Theta_{21,0} = R\cos(\psi - \delta).$$

For model $m \in \{H, O\}$, let (a_m, b_m) be its angle interval and let $L_m = b_m - a_m$. Explicitly,

$$a_H = -\gamma, \quad b_H = \pi/2 + \gamma, \quad L_H = \pi/2 + 2\gamma,$$
$$a_O = -\pi/4, \quad b_O = 3\pi/4, \quad L_O = \pi.$$

The positive-response angles end at $\delta + \pi/2$, which lies strictly inside both intervals. Their lengths give

$$P_m(\Theta_{21,0} > 0) = \frac{\pi/2 + \delta - a_m}{L_m}.$$

Thus

$$P_H(\Theta_{21,0} > 0) = 0.9667181, \qquad P_O(\Theta_{21,0} > 0) = 0.8711189.$$

Negative responses remain admissible because $b_m > \delta + \pi/2$ in both models.

For the quantiles, put $d_m = \delta - a_m$ and $e_m = b_m - \delta$. Here $0 < d_m < e_m < \pi$. For $-R \le x \le R$, let $v(x) = \arccos(x/R)$. The angles yielding a response above x have total length

$$\min\{d_m, v(x)\} + \min\{e_m, v(x)\}.$$

Consequently the CDF is

$$F_m(x) = 1 - \frac{\min\{d_m, v(x)\} + \min\{e_m, v(x)\}}{L_m}.$$

Solving $F_m(q_{p,m}) = p$ gives

$$q_{p,m} = \begin{cases} R\cos(e_m - pL_m), & p \le (e_m - d_m)/L_m, \\ R\cos\{(1-p)L_m/2\}, & p \ge (e_m - d_m)/L_m. \end{cases}$$

For both models the branch threshold lies strictly between 0.05 and 0.50. The three quantiles used in the table are therefore

$$q_{0.05,m} = R\cos(b_m - \delta - L_m/20),$$
$$q_{0.50,m} = R\cos(L_m/4),$$
$$q_{0.95,m} = R\cos(L_m/40).$$

Substitution gives

$$(q_{0.05,H}, q_{0.50,H}, q_{0.95,H}) = (0.0449674,\ 0.8737346,\ 1.0749333),$$
$$(q_{0.05,O}, q_{0.50,O}, q_{0.95,O}) = (-0.2641785,\ 0.7615773,\ 1.0737128).$$

The equal-tailed 90% interval is $[q_{0.05,m}, q_{0.95,m}]$, of width 1.0299658 under the heteroskedastic model and 1.3378913 under the homoskedastic model. The sign contrast follows from different admissible rotations in the common IRF formula. These calculations use the specified parameters and impose no estimation assumption.

Appendix S3. Complete derivation of Example 2

Let P_H and P_O denote probabilities under the respective restricted-Haar laws for both models, conditional on the specified parameters and the specified historical observation. The subscripts H and O refer to the heteroskedastic and homoskedastic models. For a positive unit shock in the first component, with the other shock set to zero, its first column is

$$\Theta_0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (A_0^{-1})' q_1 = \begin{pmatrix} \sin\psi \\ \cos\psi \end{pmatrix}.$$

Thus, $\Theta_{11,0} = \sin\psi$ is the unrestricted response of y_1, and $\Theta_{21,0} = \cos\psi$ is the response of y_2.

Step 1: Haar rotations and the sign checks. Before imposing restrictions, draw

$$Q = \begin{pmatrix} \cos\psi & -\iota\sin\psi \\ \sin\psi & \iota\cos\psi \end{pmatrix}, \qquad \psi \sim \text{Uniform}(-\pi, \pi], \qquad P(\iota = -1) = P(\iota = 1) = \tfrac{1}{2},$$

with ψ and ι independent. This gives normalized Haar measure on the full orthogonal group. The restrictions concern only the first shock, whose column is $q_1 = (\cos\psi, \sin\psi)'$.

Require shock 1 to raise y_2 in the reported responses and in each regime. In the heteroskedastic model, its regime-k impact column is

$$(A_0^{-1})' \Lambda_k^{1/2} q_1 = \begin{pmatrix} \sqrt{\lambda_{2,k}}\sin\psi \\ \sqrt{\lambda_{1,k}}\cos\psi \end{pmatrix}.$$

The three checks for the y_2 response are therefore

$$\cos\psi > 0, \qquad \sqrt{32/17}\cos\psi > 0, \qquad \sqrt{2/17}\cos\psi > 0.$$

Because both square roots are positive, these are the same condition, $\cos\psi > 0$. The homoskedastic model has the same condition. In both models, the sign-admissible angle therefore has the law

$$\psi \sim \text{Uniform}(-\pi/2, \pi/2).$$

This example retains checks in every regime. They repeat the same sign condition and exclude no additional rotations.

Step 2: Results without the narrative restriction. The response is $\Theta_{11,0} = \sin\psi$, and sine is strictly increasing on the sign-admissible interval. Its sign probability, median, and pth quantile in either model are

$$P(\Theta_{11,0} < 0) = \tfrac{1}{2}, \qquad q_{0.50} = 0, \qquad q_p = \sin(-\pi/2 + p\pi).$$

The equal-tailed 90% interval is

$$[q_{0.05}, q_{0.95}] = [-\cos(\pi/20), \cos(\pi/20)] = [-0.9876883,\ 0.9876883],$$

with width 1.9753767. Thus the entire sign-only response distribution, not just its sign probability or interval, is identical across models.

Step 3: The common historical observation. Choose the illustrative innovation $\eta_\tau = (1/4, -1)'$. It gives the observed residual

$$u_\tau = (A_0^{-1})'\eta_\tau = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1/4 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1/4 \end{pmatrix}.$$

Equivalently, starting from this observed residual recovers

$$A_0' u_\tau = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 1/4 \end{pmatrix} = \begin{pmatrix} 1/4 \\ -1 \end{pmatrix}.$$

This observation is specified for the illustration; it is not implied by the covariance matrices. Both models use the same observation. For the heteroskedastic model, date τ belongs to regime 1.

Step 4: Recovering the historical shock and imposing the narrative. The narrative requires a positive historical impact contribution of shock 1 to y_2. Under the heteroskedastic model, the recovered standardized innovation is

$$\begin{aligned} \varepsilon_\tau &= \Lambda_1^{-1/2} A_0' u_\tau \\ &= \begin{pmatrix} \sqrt{17/32}/4 \\ -\sqrt{17/2} \end{pmatrix} = \sqrt{17/2} \begin{pmatrix} 1/16 \\ -1 \end{pmatrix}. \end{aligned}$$

The recovered first shock is

$$e_{1,\tau} = q_1' \varepsilon_\tau = \sqrt{17/2}\left(\tfrac{1}{16}\cos\psi - \sin\psi\right).$$

Its regime-1 impact response in y_2 is $\sqrt{32/17}\cos\psi$, so its historical impact contribution is

$$\sqrt{32/17}\cos\psi\ e_{1,\tau}.$$

The sign checks make the first factor positive. Therefore the narrative restriction is equivalent to

$$\tfrac{1}{16}\cos\psi - \sin\psi > 0.$$

Under the homoskedastic model,

$$\varepsilon_\tau = A_0' u_\tau = (1/4, -1)', \qquad e_{1,\tau} = \tfrac{1}{4}\cos\psi - \sin\psi.$$

Its historical impact contribution to y_2 is

$$\cos\psi \; e_{1,\tau}.$$

Again the impact response is positive by the sign restriction, so the same narrative is equivalent to

$$\tfrac{1}{4}\cos\psi - \sin\psi > 0.$$

The narrative restricts the historical contribution, not the size of the shock in the response experiment. The recovered historical shock need not equal one. No dominance or variance-share condition is imposed.

Step 5: Admissible rotations after the narrative restriction. On $(-\pi/2, \pi/2)$, cosine is positive and tangent is strictly increasing. The narrative restrictions therefore become

$$\tan\psi < 1/16 \quad \text{under the heteroskedastic model,}$$
$$\tan\psi < 1/4 \quad \text{under the homoskedastic model.}$$

The resulting restricted-Haar laws are

$$\psi \sim \text{Uniform}(-\pi/2, \arctan(1/16)) \quad \text{under the heteroskedastic model,}$$
$$\psi \sim \text{Uniform}(-\pi/2, \arctan(1/4)) \quad \text{under the homoskedastic model.}$$

The sign checks alone select identical rotations across models. The different standardizations of the same historical residual make the narrative checks select different subsets.

Step 6: Sign probabilities and quantiles after the narrative restriction. For $m \in \{H, O\}$, define

$$b_H = \arctan(1/16), \qquad b_O = \arctan(1/4), \qquad L_m = \pi/2 + b_m.$$

Since $\Theta_{11,0} = \sin\psi$ is strictly increasing on each admissible interval,

$$P_m(\Theta_{11,0} < 0) = \frac{\pi/2}{L_m}, \qquad q_{p,m} = \sin\left(-\frac{\pi}{2} + pL_m\right).$$

In particular,

$$q_{0.05,m} = \sin\left(-\frac{\pi}{2} + \frac{L_m}{20}\right),$$
$$q_{0.50,m} = \sin\left(-\frac{\pi}{2} + \frac{L_m}{2}\right),$$
$$q_{0.95,m} = \sin\left(-\frac{\pi}{2} + \frac{19L_m}{20}\right).$$

Substitution gives

$$\begin{aligned} P_H(\Theta_{11,0} < 0) &= 0.9617816, \\ (q_{0.05,H}, q_{0.50,H}, q_{0.95,H}) &= (-0.9966676,\ -0.6846976,\ -0.0192408), \\ P_O(\Theta_{11,0} < 0) &= 0.8650831, \\ (q_{0.05,O}, q_{0.50,O}, q_{0.95,O}) &= (-0.9958815,\ -0.6154122,\ 0.1535797). \end{aligned}$$

The equal-tailed interval widths are 0.9774269 and 1.1494612, respectively. The table rounds entries to four decimal places, using unrounded endpoints for widths. Positive responses remain admissible because both upper angle bounds are positive. The heteroskedastic negative-sign probability exceeds 0.95, whereas the homoskedastic probability is below 0.95; this is why only the former has a wholly negative equal-tailed 90% interval.

Appendix S4. Invariance of Algorithm 1

In this section, we show that Algorithm 1 preserves the restricted conditional target. Fix Q and s_0. Suppress s_0 where no ambiguity arises and write

$$f(d\theta) = p_0(dz_0 \mid s_0)p_+(dz_+ \mid z_0, s_0).$$

Let $K_0(a, da')$ be the first-block MH transition, including its rejection outcome. It satisfies

$$p_0(da \mid s_0)K_0(a, da') = p_0(da' \mid s_0)K_0(a', da).$$

For completed states $x = (a, b)$ and $y = (a', b')$, the MH-then-companion transition is

$$C(x, dy) = K_0(a, da')p_+(db' \mid a', s_0).$$

It is reversible for f, since

$$\begin{aligned} f(dx)C(x, dy) &= p_0(da \mid s_0)K_0(a, da')p_+(db \mid a, s_0)p_+(db' \mid a', s_0) \\ &= p_0(da' \mid s_0)K_0(a', da)p_+(db' \mid a', s_0)p_+(db \mid a, s_0) \\ &= f(dy)C(y, dx). \end{aligned}$$

Let

$$S = \{\theta : SR(\theta, Q) = 1\}, \qquad Z = f(S) > 0, \qquad \pi_S(d\theta) = \frac{\mathbf{1}_S(\theta)f(d\theta)}{Z}.$$

Starting at $x_0 \in S$, generate x_κ from $C(x_{\kappa-1}, dx_\kappa)$ and let

$$\tau = \inf\{\kappa \geq 1 : x_\kappa \in S\}, \qquad N = N_\theta + 1.$$

The algorithm returns x_τ if $\tau \leq N$ and returns x_0 otherwise. The budget N is fixed before the path is generated and is the same for forward and reverse moves.

For any $\nu \leq N$, a first-return path has $x_0, x_\nu \in S$ and $x_1, \ldots, x_{\nu-1} \notin S$. Reversibility of C implies equality of the forward and reverse path measures:

$$\begin{aligned} &f(dx_0)C(x_0, dx_1)\cdots C(x_{\nu-1}, dx_\nu) \\ &\quad = f(dx_\nu)C(x_\nu, dx_{\nu-1})\cdots C(x_1, dx_0). \end{aligned}$$

The reversed path has the same endpoints in S, the same intermediate states outside S, and the same length ν. Restricting to these paths, dividing by Z, integrating out intermediate states, and summing over $\nu \leq N$ gives detailed balance for every move between distinct admissible endpoints. Paths with no return by N contribute only to the probability of retaining x_0. Thus the finite-budget transition is reversible for π_S. No endpoint acceptance probability is needed.

This proves invariance of Steps 1–2 for the parameter conditional required by Definition 1. The companion refresh in Step 3, the stated latent full-conditional updates, and the ESS rotation updates preserve their respective conditionals. Their composition therefore preserves the joint target. Convergence additionally requires that the full chain can explore that target.

For a simple finite illustration, take a uniform unrestricted target on a, b, c and

$$C = \begin{pmatrix} 0.6 & 0.1 & 0.3 \\ 0.1 & 0.8 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}, \qquad S = \{a, b\}.$$

With a budget of two transitions, continuing from the failed state c gives

$$P(a, b) = 0.1 + 0.3(0.1) = 0.13, \qquad P(b, a) = 0.1 + 0.1(0.3) = 0.13.$$

These moves balance the uniform restricted target. Resetting to the original state after failure instead gives 0.13 and 0.11. The difference is the continuation rule, not the order of two tests on an otherwise unchanged proposal.

Appendix S5. Data

Most data series are from Federal Reserve Economic Data (FRED), the database of the Federal Reserve Bank of St. Louis. We report FRED series IDs whenever available.

Monetary application.

- Real output: `GDPC1`. The series is interpolated from quarterly to monthly frequency.
- GDP deflator: `GDPDEF`. The series is interpolated from quarterly to monthly frequency.
- Commodity price index: `PSCCOM` (Haver Analytics; not available on FRED).
- Total reserves: `TRARR`.
- Non-borrowed reserves: `BOGNONBR`.
- Federal funds rate: `FEDFUNDS`.

Oil market application.

- Oil production: world crude oil production, U.S. Energy Information Administration (EIA) Monthly Energy Review; not available on FRED.
- Real global activity: Kilian's index of global real economic activity, from the ADRR replication files (`Kilian_Data_Updated.mat`), the same series CMT use.

• Real oil price: U.S. refiners' acquisition cost of imported crude oil, from the EIA; not available on FRED. The series is deflated by the CPI, `CPIAUCSL`.

Fiscal application.

• Net taxes: total government current receipts, `GRECPT`, less government current transfer payments, `A084RC1Q027SBEA`, and government interest payments, `A180RC1Q027SBEA`.

• Government spending: `GCE`.

• GDP: `GDP`.

• Treasury bill rate: `TB3MS`, averaged from monthly to quarterly frequency.

• GDP deflator: `GDPDEF`.

• Population: National Income and Product Accounts (NIPA) midperiod population, `B230RC0Q173SBEA`.

Appendix S6. Computational Details

This appendix provides supporting material for Section VIII. We now report the hardware and software environment used for the timing results: all computations use MATLAB R2025a on a desktop workstation with an Intel Core i9-14900F CPU (24 cores; base frequency 2.0 GHz). The remainder of the appendix reports the full HARS–CMT impulse-response comparisons and the per-application timing tables.

S6.1. **CMT comparison and supplementary impulse responses.** We use the same data, lag specifications, variance regimes, and economic restrictions. The two approaches retain their respective posterior targets, so the comparison concerns both their inferential frameworks and their computational implementations. The HARS panels use 100,000 saved draws. The CMT panels use fewer saved draws because its accept–reject step is costly: 500 in the monetary application and 1,000 in the oil application. HARS uses Student-t shocks, while these CMT runs use Gaussian shocks. The short CMT runs illustrate the comparison; their tail endpoints do not establish the main economic conclusions.

Monetary policy application. Figure S1 compares HARS with CMT's Algorithm 2 for the contractionary monetary policy shock. The two samplers agree on the sign and timing of the output contraction. They differ more clearly on the GDP deflator and total reserves. Under CMT, the deflator decline is weak once the imposed sign horizon ends, and the response of total reserves is not tightly negative. HARS instead delivers a persistent deflator decline and a reserve contraction over roughly the first year. These differences are the main reason the selected comparison appears in Figure 5 of the main text.

Oil market application. Figure S2 compares the two samplers for the three oil market shocks. The main difference is the response of real economic activity to an oil-specific demand shock. HARS recovers a short-horizon contraction. Under CMT, the 90% band leaves the contraction uncertain at the unrestricted early horizons. The impact response has

the imposed negative sign. This is the second selected comparison summarized in Figure 5 of the main text.

Fiscal shock application. We do not give a fiscal CMT impulse-response comparison. The fiscal application identifies the net-tax and government-spending shocks simultaneously and imposes the narrative restrictions on both shocks. In our implementation with both shocks restricted simultaneously, CMT's Algorithm 2 accepts no draws among 200,000 candidates. This run therefore supplies no posterior bands.

S6.2. **Timing and repair details.** Tables S1, S2, and S3 compare the time needed to obtain 1,000 effective impulse response draws. The HARS entries use the repair-enabled ($N_\theta = 100$) runs with 100,000 saved draws of (θ, Q) in Table 10. The same runs supply the HARS panels of the comparison figures in Section S6.1. We compute effective sample size with the multivariate estimator of Vats et al. (2019) over the impulse responses at horizons 0–11 in every application. All runs use the same hardware. Table 10 also compares HARS with repair disabled ($N_\theta = 0$).

Monetary policy application. Table S1 gives 6.60 hours for CMT's Algorithm 2 and 0.097 hours for HARS per 1,000 effective impulse response draws, a ratio of about 68 (see also Table S4). The CMT timing comes from a short 500-draw run with 10 burn-in iterations, and the ratio indicates the order of magnitude.

Oil market application. Table S2 gives 1.095 hours for CMT's Algorithm 2 and 0.034 hours for HARS per 1,000 effective impulse response draws, a ratio of about 32.

Fiscal shock application. Table S3 gives 0.035 hours for HARS per 1,000 effective impulse response draws. CMT's Algorithm 2 accepts no draws among 200,000 candidates in this run; a timing for effective draws cannot be computed.

S6.3. **Sampler settings and Monte Carlo precision.** This section gives the sampler settings, shock distributions, Monte Carlo standard errors, and multi-chain diagnostics from our runs.

Settings. Every HARS run in the paper uses $N_{\text{ess}} = 5$ rotation updates per parameter iteration, 3,000 burn-in iterations, and 20,000 parameter iterations after burn-in. These give the 100,000 saved pairs of (θ, Q), with no thinning. The first-block proposal q_0 is a Gaussian random walk with covariance cH, where H is the inverse Hessian at the posterior mode of (A_0, Λ). During the first 2,000 burn-in iterations, c adapts toward an acceptance rate of 0.30. We then freeze c at its geometric mean over iterations 1,001 to 2,000. The frozen c stays fixed for the rest of the run, including every repair step. It lies between 0.22 and 0.30 in the heteroskedastic runs and equals 0.78 in the homoskedastic fiscal run. The heteroskedastic runs have the Student-t shocks of Section II.1. The homoskedastic runs, ADRR, KM, and CMT have Gaussian shocks.

Efficiency. The multivariate estimator of Vats et al. (2019) uses batch means with batch size $\lfloor N^{1/4} \rfloor$. The response vector stacks horizons 0 to 11 of all variables and identified shocks. Before any constant coordinates are omitted, its dimension is 72 (monetary), 108 (oil), and 120 (fiscal). The determinant-based estimator requires nonsingular covariance estimates; coordinates fixed by a shock normalization must be omitted. In every HARS run, the estimated effective sample size lies between 10% and 19% of the saved draws.

Monte Carlo precision. We compute Monte Carlo standard errors from the saved draws behind the reported results, using 50 batches of 400 consecutive parameter iterations. Each batch keeps the five rotations of an iteration together. The standard errors of the posterior probabilities in Table 3 are at most 0.008.

Multiple chains. In the fiscal application, four chains each at $N_\theta = 100$ and $N_\theta = 200$ start from the two regions explored without repair and from two dispersed points. The pooled bands at $N_\theta = 100$ and $N_\theta = 200$ differ by at most 3% of band width and have identical sign intervals. In the monetary and oil applications, each reported result comes from one chain.

TABLE S1. Sampling efficiency, monetary application

Sampler	**Time per** 1,000 **effective IRF draws (hours)**
CMT, heteroskedasticity	6.60
HARS, heteroskedasticity	0.097

Notes: Effective sample size is computed over the impulse responses of the six variables at horizons 0–11 to the monetary policy shock. Saved draws after burn-in: 100,000 draws of (θ, Q) for HARS and 500 for CMT. The HARS run enables posterior-draw repair ($N_\theta = 100$). The CMT entry comes from this short run, with 10 burn-in iterations, and indicates the order of magnitude.

TABLE S2. Sampling efficiency, oil market application

Sampler	**Time per** 1,000 **effective IRF draws (hours)**
CMT, heteroskedasticity	1.095
HARS, heteroskedasticity	0.034

Notes: Effective sample size is computed over the impulse responses of the three variables at horizons 0–11 to all three identified shocks. Saved draws after burn-in: 100,000 draws of (θ, Q) for HARS and 1,000 for CMT. The CMT entry uses the full 1,000-draw run. The HARS run enables posterior-draw repair ($N_\theta = 100$); only (A_0, Λ) enters the oil admissibility check because the restrictions bind only on impact.

TABLE S3. Sampling efficiency, fiscal application

Sampler	**Time per** 1,000 **effective IRF draws (hours)**
CMT, heteroskedasticity	*fails*
HARS, heteroskedasticity	0.035

Notes: Effective sample size is computed over the impulse responses of the five variables at horizons 0–11 to the two fiscal shocks identified simultaneously. Saved draws after burn-in: 100,000 draws of (θ, Q) for HARS, with posterior-draw repair enabled ($N_\theta = 100$). The run uses the 3-regime heteroskedastic specification. Under the same sign, narrative, and 3-regime restrictions, with both fiscal shocks identified simultaneously, our implementation of CMT's Algorithm 2 accepts no draws among 200,000 candidates. Its time per effective draw cannot be computed.

TABLE S4. Computing time: CMT vs. HARS

Application	**CMT** (hours)	**HARS** (hours)	**Speed-up** (HARS)
Monetary	6.60	0.097	$\approx 68\times$
Oil	1.095	0.034	$\approx 32\times$
Fiscal	*fails*	0.035	—

Notes: Hours per 1,000 effective impulse response draws, computed as in Tables S1, S2, and S3. The HARS entries are the repair-enabled ($N_\theta = 100$) timings from Table 10. All runs use the same hardware (Appendix S6). The monetary CMT entry comes from a short 500-draw run, and its speed ratio indicates the order of magnitude. In the fiscal application, both shocks are identified simultaneously under the same sign, narrative, and 3-regime restrictions; *fails* denotes no admissible draws among 200,000 candidates in our CMT implementation. Timing ratios compare these implementations under their respective posterior targets.

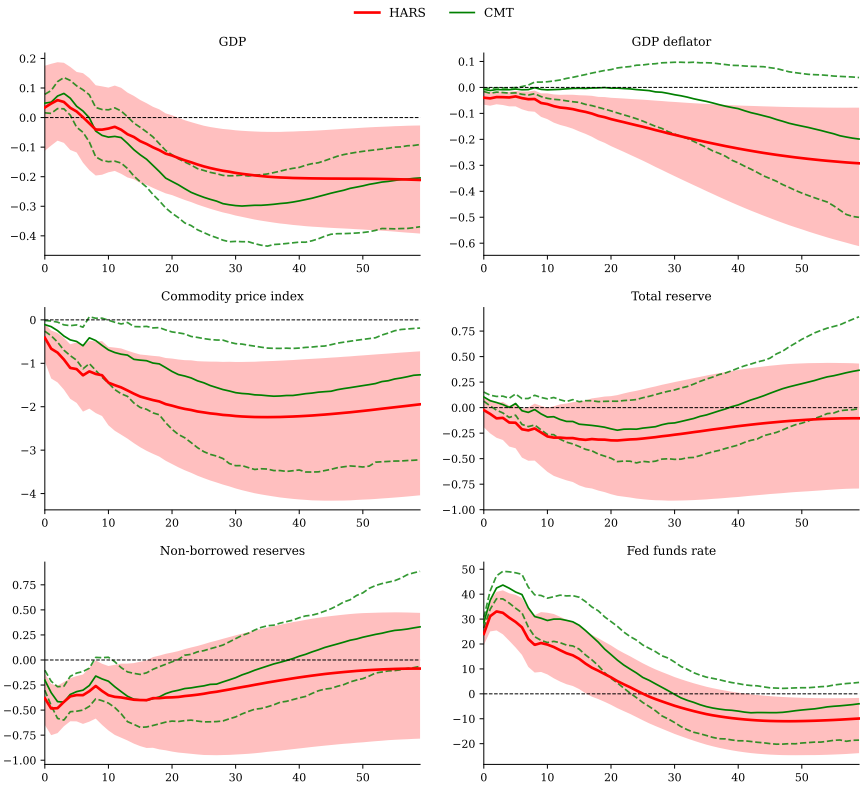


FIGURE S1. Impulse responses to a contractionary monetary policy shock: HARS versus CMT's Algorithm 2, both under the 4-regime specification. Solid lines are posterior medians and bands are 90% credible intervals. Vertical axes are in percent, except for the federal funds rate, which is in basis points.

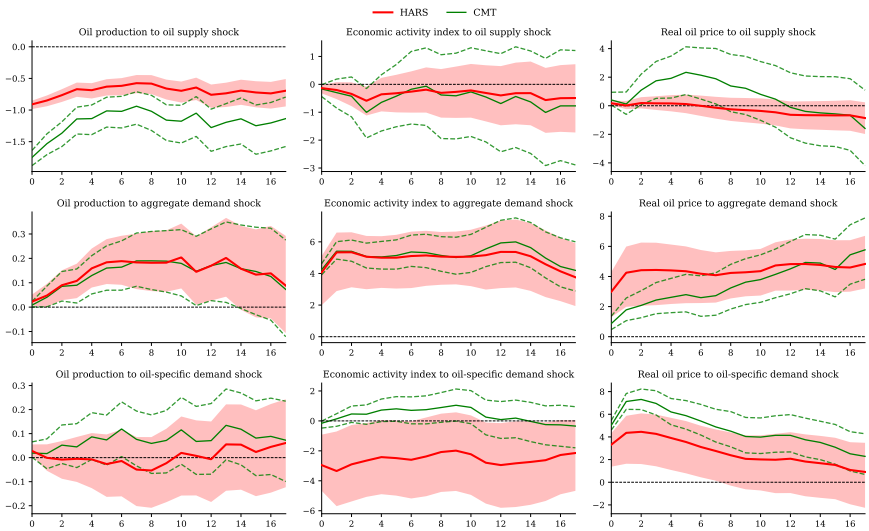


FIGURE S2. Impulse responses to the three identified oil market shocks: HARS versus CMT's Algorithm 2, both under the 3-regime specification. Solid lines are posterior medians and bands are 90% credible intervals.

References

Juan Antolín-Díaz and Juan F. Rubio-Ramírez. Narrative sign restrictions for SVARs. *American Economic Review*, 108(10):2802–2829, 2018.

Jonas E. Arias, Juan F. Rubio-Ramírez, and Daniel F. Waggoner. Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica*, 86(2):685–720, 2018.

Jonas E. Arias, Juan F. Rubio-Ramírez, and Minchul Shin. Are fiscal transfers inflationary? Working paper, Federal Reserve Bank of Philadelphia, February 2026a.

Jonas E. Arias, Juan F. Rubio-Ramírez, and Minchul Shin. Large SVARs. Federal Reserve Bank of Philadelphia Working Paper 26-04, January 2026b.

Jonas E. Arias, Juan F. Rubio-Ramírez, Minchul Shin, and Daniel F. Waggoner. Inference based on time-varying SVARs identified with sign restrictions. *Review of Economic Studies*, 00(0):1–30, 2026c. doi: 10.1093/restud/rdag008. Advance Access publication January 2026.

Michael T. Belongia and Peter N. Ireland. The transmission of monetary policy shocks through the markets for reserves and money. *Journal of Macroeconomics*, 80:103590, 2024.

Francesco Bianchi, Martin Lettau, and Sydney C. Ludvigson. Monetary policy and asset valuation. *The Journal of Finance*, 77(2):967–1017, 2022. doi: 10.1111/jofi.13107.

Olivier Blanchard and Roberto Perotti. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*, 117(4):1329–1368, 2002.

Markus K. Brunnermeier, Darius Palia, Karthik A. Sastry, and Christopher A. Sims. Feedbacks: Financial markets and economic activity. *American Economic Review*, 111:1845–1879, 2021. DOI: 10.1257/aer.20180733.

Dario Caldara and Christophe Kamps. The analytics of SVARs: a unified framework to measure fiscal multipliers. *The Review of Economic Studies*, 84(3):1015–1040, 2017.

Andrea Carriero, Massimiliano Marcellino, and Tommaso Tornese. Blended identification in structural VARs. *Journal of Monetary Economics*, 146:103581, 2024. doi: 10.1016/j.jmoneco.2024.103581.

Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. The effects of monetary policy shocks: Evidence from the flow of funds. *Review of Economics and Statistics*, 78 (1):16–34, 1996.

Helmut Herwartz and Helmut Lütkepohl. Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks. *Journal of Econometrics*, 183(1):104–116, 2014. doi: 10.1016/j.jeconom.2014.06.011.

Peter L. Jørgensen and Søren Hove Ravn. The inflation response to government spending shocks: a fiscal price puzzle? *European Economic Review*, 141:103982, 2022.

Lutz Kilian and Daniel P. Murphy. Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models. *Journal of the European Economic Association*, 10(5):1166–1188, 2012.

Lutz Kilian and Kunal Patel. What the Iran War teaches us about the price elasticity of oil supply. Working Paper 2625, Federal Reserve Bank of Dallas, August 2026. URL https://www.dallasfed.org/research/papers/2026/wp2625.

Karel Mertens and Morten O. Ravn. The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review*, 103(4):1212–1247, 2013.

Andrew Mountford and Harald Uhlig. What are the effects of fiscal policy shocks? *Journal of Applied Econometrics*, 24(6):960–992, 2009.

Iain Murray, Ryan P. Adams, and David J. C. MacKay. Elliptical slice sampling. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9 of *Proceedings of Machine Learning Research*, pages 541–548, Chia Laguna Resort, Sardinia, Italy, 2010. PMLR.

Roberto Perotti. Estimating the effects of fiscal policy in OECD countries. CEPR Discussion Paper 4842, Centre for Economic Policy Research, 2005.

Giorgio E. Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005. doi: 10.1111/j.1467-937X.2005.00353.x.

Valerie A. Ramey. Identifying government spending shocks: It's all in the timing. *Quarterly Journal of Economics*, 126(1):1–50, 2011.

Valerie A. Ramey. Macroeconomic shocks and their propagation. In John B. Taylor and Harald Uhlig, editors, *Handbook of Macroeconomics*, volume 2, chapter 2, pages 71–162. Elsevier, 2016. doi: 10.1016/bs.hesmac.2016.03.003.

Christina D. Romer and David H. Romer. The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, 100(3):763–801, 2010. doi: 10.1257/aer.100.3.763.

Juan F. Rubio-Ramírez, Daniel F. Waggoner, and Tao Zha. Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77(2):665–696, 2010. doi: 10.1111/j.1467-937X.2009.00578.x.

Christopher A. Sims and Tao Zha. Error bands for impulse responses. *Econometrica*, 67(5):1113–1155, 1999.

Christopher A. Sims and Tao Zha. Were there regime switches in u.s. monetary policy? *American Economic Review*, 96:54–81, 2006.

Harald Uhlig. What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419, 2005.

Dootika Vats, James M. Flegal, and Galin L. Jones. Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106(2):321–337, 2019. doi: 10.1093/biomet/asz002.

Daniel F. Waggoner and Tao Zha. A gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control*, 28(2):349–366, 2003.